

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:00:47 ; Search time 45 Seconds
(without alignments)
704,748 Million cell updates/sec

Title: US-09-967-301-2

Perfect score: 1275
Sequence: 1 MSKGEELFTGVVPIVLVDG.....VLLEFVTAAAGITHCMDELYK 238

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1275	100.0	238	17 AAW05304	Green fluorescent
2	1275	100.0	238	18 AAW24232	Aequorea victoria
3	1275	100.0	238	19 AAW76371	A. victoria green
4	1275	100.0	238	19 AAW76105	A. victoria green
5	1275	100.0	238	19 AAW40479	A. victoria green
6	1275	100.0	238	22 AAB73552	Wild-type green fl
7	1275	100.0	238	23 AAE16038	Aequorea victoria
8	1274	99.9	238	19 AAW65084	A. victoria green
9	1272	99.8	238	19 AAW76109	A. victoria green
10	1272	99.8	238	19 AAW76110	A. victoria green

11	1272	99.8	238	19 AAW76111	A. victoria green
12	1272	99.8	238	19 AAW5079	A. victoria green
13	1272	99.8	238	19 AAW5079	A. victoria green
14	1272	99.8	238	23 AAE16042	Aequorea victoria
15	1272	99.8	238	23 AAE16043	Aequorea victoria
16	1272	99.8	501	18 AAW31879	GFP variants S65T
17	1271	99.7	238	17 AAW05312	Green fluorescent
18	1271	99.7	238	18 AAW36057	Green fluorescent
19	1271	99.7	238	18 AAW31295	Aequorea victoria
20	1271	99.7	238	19 AAW76113	A. victoria green
21	1271	99.7	238	19 AAW65081	A. victoria green
22	1271	99.7	238	19 AAW65085	A. victoria green
23	1271	99.7	238	20 AAY43449	Green fluorescent
24	1271	99.7	238	20 AAW88484	Green fluorescent
25	1271	99.7	238	21 AAB01016	Aequorea victoria
26	1271	99.7	238	21 AAB03411	Aequorea victoria
27	1271	99.7	238	21 AAB15661	Humanised GFP, Ho
28	1271	99.7	238	21 AAB01608	Aequorea victoria
29	1271	99.7	238	21 AAY99804	Aequorea victoria
30	1271	99.7	238	21 AAY99835	Aequorea victoria
31	1271	99.7	238	21 AAY99723	Aequorea victoria
32	1271	99.7	238	21 AAY99732	Aequorea victoria
33	1271	99.7	238	21 AAY70365	Humanised green fl
34	1271	99.7	238	22 AAB73554	Green fluorescent
35	1271	99.7	238	22 AAB85899	A. victoria green
36	1271	99.7	238	22 AAG66199	A. victoria green
37	1271	99.7	238	23 AAE14600	Aequorea victoria
38	1271	99.7	238	23 AAE16041	Aequorea victoria
39	1271	99.7	238	23 AAE16056	Aequorea victoria
40	1271	99.7	239	21 AAY59283	Amino acid sequenc
41	1270	99.6	238	17 AAW05306	Green fluorescent
42	1270	99.6	238	19 AAW59822	Green fluorescent
43	1270	99.6	238	19 AAW65086	A. victoria green
44	1270	99.6	238	19 AAW65082	A. victoria green
45	1270	99.6	238	22 AAB73555	Green fluorescent

ALIGNMENTS

RESULT 1
AAW05304
ID AAW05304 standard; Protein: 238 AA.
XX AAW05304:
AC AAW05304:
XX
DT 02-APR-1997 (first entry)
XX
DE Green fluorescent protein.
XX
KW Green fluorescent protein; GFP; jellyfish; Aequorea victoria; sea pansy;
KW Renilla reniformis; differential gene expression; protein localisation;
KW gene expression tracking; fluorescence.
XX
OS Aequorea victoria.
XX
PN W09623810-A1
XX
PD 08-AUG-1996.
XX
PF 13-NOV-1995; 95WO-US14692.
XX
PR 10-NOV-1994; 94US-0337915.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Heim R, Tsien RY;
XX
DR WPI: 1996-371370/37.
XX
PT N-PSDB: AAT39694.
XX
PT New modified Aequorea green fluorescent polypeptide(s) - having amino acid changes to provide prods. which exhibit different

PT excitation and emission spectra
 XX
 PS
 XX Disclosure; Page 20-21; 39pp; English.
 XX
 CC This sequence represents the wild type Aequorea victoria (jellyfish)
 CC green fluorescent protein (GFP). The fluorescence of this protein is
 CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
 CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
 CC the one absorbance peak seen in the related GFP from the sea pansy
 CC (Renilla reniformis). This protein is modified, to provide the GFPs of
 CC the invention. The modifications were created by subjecting the cDNA
 CC encoding this sequence to site directed mutagenesis using mutagenic PCR
 CC primers, or hydroxylamine treatment. The GFPs of the invention are
 CC modified to lead to the formation of products with markedly different
 CC excitation and emission spectra. Visibly distinct colours, and increased
 CC intensities of emission make these products useful in a wide variety of
 CC contexts, such as tracking of differential gene expression and protein
 CC localisation. The mutations can also be created to modify the encoded
 CC GFP so that it only possesses one absorbance peak.
 XX
 SQ Sequence 238 AA;
 Query Match 100.0%; Score 1275; PB 17; Length 238;
 Best Local Similarity 100.0%; Pred. No. 4.6e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTFLKFICTTGKLPVNPPTL 60
 DB 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTFLKFICTTGKLPVNPPTL 60
 QY 61 VTTFSGVQCFSRYPDHMKRHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTFSGVQCFSRYPDHMKRHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRTELKGIKDEKDNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVSQVLAD 180
 DB 121 NRTELKGIKDEKDNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVSQVLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYK 238
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYK 238
 RESULT 2
 AAW24232
 ID AAW24232 standard; Protein; 238 AA.
 XX
 AC AAW24232;
 XX
 DT 02-FEB-1998 (first entry)
 XX
 DE Aequorea victoria green fluorescent protein (GFP).
 XX
 KW North West Pacific jellyfish; green fluorescent protein; GFP; S65C;
 KW fluorescent resonance energy transfer; FRET; enzymatic assay;
 KW enzyme inhibitor; enzyme promoter; Tandem fluorescent protein construct;
 KW blue fluorescent protein; protease cleavage site.
 XX
 OS Aequorea victoria.
 XX
 PN WO9728261-A1.
 XX
 PD 07-AUG-1997.
 XX
 PF 31-JAN-1997; 97WO-US01457.
 XX
 PP 31-JAN-1996; 96US-0594575.
 XX
 PR (AURO-) AURORA BIOSCIENCES CORP.
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Cubitt A. Heim R. Tsien RY;
 XX

DR WPI; 1997-402615/37.
 DR N-PSDR; AAT77843.
 XX
 PT Tandem fluorescent protein constructs - have donor and acceptor
 PT moieties exhibiting fluorescent linked via cleavable peptide linker.
 PT useful in enzymatic assays
 XX
 PS Disclosure; Fig 1; 88pp; English.
 CC
 CC This sequence represents Aequorea victoria (North West Pacific jellyfish)
 CC green fluorescent protein (GFP). The GFP protein is used in a novel
 CC tandem fluorescent protein construct. The construct comprises a donor
 CC (e.g. GFP protein) and an acceptor (e.g. S65C, a variant of GFP or
 CC blue fluorescent protein) fluorescent protein moiety, and a linker
 CC coupling them. Preferably, the donor is positioned at the N-terminus of
 CC the polypeptide relative to the acceptor. The linker moiety is a peptide
 CC 5-50 amino acids in length containing a protease cleavage site. The donor
 CC and acceptor moieties exhibit fluorescent resonance energy transfer
 CC (FRET) when the donor moiety is excited. The constructs are used in
 CC enzymatic assays and can be used to isolate new enzymes or enzyme
 CC inhibitors or promoters. The specific activity of enzyme (in vivo and in
 CC vitro) and compounds altering enzyme activity can be obtained. FRET and
 CC hence activity of specific compounds is measured from the acceptor or
 CC donor moiety or maybe obtained using a ratio between the two.
 XX
 SQ Sequence 238 AA;
 Query Match 100.0%; Score 1275; DB 18; Length 238;
 Best Local Similarity 100.0%; Pred. No. 4.6e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTFLKFICTTGKLPVNPPTL 60
 DB 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGKLTFLKFICTTGKLPVNPPTL 60
 QY 61 VTTFSGVQCFSRYPDHMKRHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTFSGVQCFSRYPDHMKRHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRTELKGIKDEKDNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVSQVLAD 180
 DB 121 NRTELKGIKDEKDNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVSQVLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYK 238
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYK 238
 RESULT 3
 AAW76371
 ID AAW76371 standard; Protein; 238 AA.
 XX
 AC AAW76371;
 XX
 DT 03-DEC-1998 (first entry)
 XX
 DE A. victoria green fluorescent protein.
 XX
 KW Green fluorescent protein; GFP; jellyfish; detection; mutant; activity;
 KW distribution.
 XX
 OS Aequorea victoria.
 XX
 PN JP10234382-A.
 XX
 PD 08-SEP-1998.
 XX
 PF 27-FEB-1997; 97JP-0062370.
 XX
 PP 27-FEB-1997; 97JP-0062370.
 XX
 PR (DINA-) DINABEKU KENKYUSHO KK.
 PA
 XX

RESULT 3
AAE16073
ID AAE16073 standard: Protein: 238 AA.
XX AC AAE16073;
XX AC AAE16073;
XX 26-MAR-2002 (first entry)
XX Aequorea victoria GFP mutant protein (E222X).
XX
XX Green fluorescent protein: cell lineage tracer; protein localisation:
KW GFP: fusion tag; gene expression marker; fluorescent energy acceptor;
KW immunocassay; hybridisation assay; fluorescent energy donor; biosensor;
KW FRET; fluorescence resonance energy transfer; mutant; mutein.
XX
XX Aequorea victoria.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 222 /label= Asn, Glu
FT /note= "Wild type Glu substituted with Xaa"
FT
XX WC200190147-A2.
XX
XX 29-NOV-2001.
XX
XX 17-MAY-2001: 2001WO-US16149.
XX
XX 19-MAY-2000: 2000US-0575847.
XX (UYOR-) UNIV OREGON STATE.
XX
XX Wachter R, Remington SJ;
XX WPI: 2002-083084/11.
XX
XX New long wavelength engineered fluorescent proteins, useful as markers
PT for gene expression, tracers of cell lineage or as fusion tags to
PT monitor protein localization, or in detection assays, e.g. immunoassays
PT or hybridization assays .
XX
XX Claim 47: Page -: 18lpp: English.
XX
XX The patent discloses functional engineered fluorescent protein and its
XX corresponding polynucleotide. The amino acid sequence of the engineered
XX protein is identical to Aequorea green fluorescent protein (GFP). The
XX engineered fluorescent proteins of the invention have varied fluorescent
XX properties and have the ability to respond to ion concentrations via a
XX change in fluorescent characteristics. They are useful as markers for
XX gene expression, tracers of cell lineage or as fusion tags to monitor
XX protein localisation within living cells. The engineered fluorescent
XX proteins are particularly useful for coupling engineered fluorescent
XX proteins to antibodies, nucleic acids or other receptors for use in
XX detection assays, e.g. immunoassays or hybridisation assays. They are
XX useful for tracking the movement of proteins in cells or in systems
XX for detecting induction of transcription and for the simultaneous
XX measurement of two or more processes within cells. Proteins of the
XX invention are also useful as fluorescent energy donors or acceptors
XX as well as biosensors for detecting anions. They are also useful in
XX fluorescence resonance energy transfer (FRET). The crystal structure
XX of the GFP is useful for designing mutants having altered fluorescent
XX characteristics which are particularly useful to identify amino acids
XX whose substitution alters fluorescent properties of the protein. The
XX crystal structure of the GFP is also useful for designing mutants having
XX altered anion binding characteristics which are particularly useful for
XX identifying amino acids whose substitution alters the specificity and
XX affinity of the binding site to various anions, and for monitoring anion
XX binding and therefore the concentration of the anion. The present
XX sequence is Aequorea victoria GFP mutant protein (E222X).
XX Note: This sequence is not shown in the specification but is derived
XX from Aequorea victoria wild type green fluorescent protein shown in

CC figure 3 of the specification (AAE16073).

XX Sequence 238 AA;

Query Match 98.7%; Score 1259; DB 23; Length 238;
Best Local Similarity 98.7%; Pred. No. 1.5e-123;
Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNGHFKFSVSGEGSDATYKGLTLKFLICITGKLPVWPMTL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNGHFKFSVSGEGSDATYKGLTLKFLICITGKLPVWPMTL 60
QY 61 VTLSYGVQCFSRYPDMKRHRDFKSAPEGYVQEMTIFFKDGNKYKTRAEYKFGDTLY 120
DB 61 VTLSYGVQCFSRYPDMKRHRDFKSAPEGYVQEMTIFFKDGNKYKTRAEYKFGDTLY 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNCIKVNEKIRHNIEDGSGVLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNCIKVNEKIRHNIEDGSGVLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLGFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLGFVTAAGITHGMDELYK 238

RESULT 4
AAW05304
ID AAW05304 standard: Protein: 238 AA.

XX AC AAW05304;

DT 02-APR-1997 (first entry)

XX Green fluorescent protein.

XX Green fluorescent protein: GFP; jellyfish; Aequorea victoria; sea pansy;
KW Renilla reniformis; differential gene expression; protein localisation;
KW gene expression tracking; fluorescence.

OS Aequorea victoria.

XX W09623810-A1.

PD 08-AUG-1996.

XX 13-NOV-1995: 95WO-US24692.

XX 10-NOV-1994: 94US-G337915.

XX (REGC) UNIV CALIFORNIA.

XX Heim R, Tsien RY;

XX WPI: 1996-371370/37.

XX N-PSDB: AAI39694.

XX New modified Aequorea green fluorescent polypeptide(s) - having
PT amino acid changes to provide prods. which exhibit different
PT excitation and emission spectra

XX Dislosure: Page 20-21; 39pp: English.

XX This sequence represents the wild type Aequorea victoria (jellyfish)
CC green fluorescent protein (GFP). The fluorescence of this protein is
CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
CC the one absorbance peak seen in the related GFP from the sea pansy
CC (Renilla reniformis). This protein is modified, to provide the GFPs of
CC the invention. The modifications were created by subjecting the cDNA
CC encoding this sequence to site directed mutagenesis using mutagenic PCR
CC primers, or hydroxylamine treatment. The GFPs of the invention are
CC modified to lead to the formation of products with markedly different
CC excitation and emission spectra. Visibly distinct colours, and increased

CC intensities of emission make these products useful in a wide variety of
 CC contexts, such as tracking of differential gene expression and protein
 CC localisation. The mutations can also be created to modify the encoded
 CC GFP so that it only possesses one absorbance peak.

XX
 SQ Sequence 238 AA;
 Query Match 98.6%; Score 1258; DB 17; Length 238;
 Best Local Similarity 98.7%; Pred. NO. 2e-123;
 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYGKILKFICTTGKLPVPMPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYGKILKFICTTGKLPVPMPTL 60
 QY 61 VTLSYGVQCFSRYPDHMKRHDFEFSKAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTLSYGVQCFSRYPDHMKRHDFEFSKAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRTELKGIQDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKVNFKIRHNIEDGGVQLAD 180
 DB 121 NRTELKGIQDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKVNFKIRHNIEDGGVQLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLGFVTAAGITHGMDELYK 238
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLGFVTAAGITHGMDELYK 238

RESULT 5
 ANW24232
 ID AAW24232 standard; Protein: 238 AA.
 XX AC AAW24232;
 XX DI 02-FEB-1998 (first entry)
 XX DE Aequorea victoria green fluorescent protein (GFP).
 XX KW North West Pacific jellyfish; green fluorescent protein; GFP; S65C;
 KW fluorescent resonance energy transfer; FRET; enzymatic assay;
 KW enzyme inhibitor; enzyme promoter; Tandem fluorescent protein construct;
 KW blue fluorescent protein; protease cleavage site.
 XX AC Aequorea victoria.
 XX PN WO9728261-A1.
 XX PD 07-AUG-1997.
 XX PF 31-JAN-1997; 97WO-US01457.
 XX PR 31-JAN-1996; 96US-0594575.
 XX PA (AURO-) AURORA BIOSCIENCES CORP.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Cubitt A. Heim R. Tsien RY;
 XX PI WP: 1997-402615/37.
 XX DR N-PSDB; AAI77843.
 XX PT Tandem fluorescent protein constructs - have donor and acceptor
 PT moieties exhibiting fluorescent linked via cleavable peptide linker,
 PT useful in enzymatic assays
 XX PS Disclosure; Fig 1; 80pp; English.

XX This sequence represents Aequorea victoria (North West Pacific jellyfish)
 CC green fluorescent protein (GFP). The GFP protein is used in a novel
 CC tandem fluorescent protein construct. The construct comprises a donor
 CC (e.g. GFP protein) and an acceptor (e.g. S65C, a variant of GFP or
 CC blue fluorescent protein) fluorescent protein moiety, and a linker
 CC coupling them. Preferably, the donor is positioned at the N-terminus of

CC the polypeptide relative to the acceptor. The linker moiety is a peptide
 CC 5-50 amino acids in length containing a protease cleavage site. The donor
 CC and acceptor moieties exhibit fluorescent resonance energy transfer
 CC (FRET) when the donor moiety is excited. The constructs are used in
 CC enzymatic assays and can be used to isolate new enzymes or enzyme
 CC inhibitors or promoters. The specific activity of enzyme (in vivo and in
 CC vitro) and compounds altering enzyme activity can be obtained. FRET and
 CC hence activity of specific compounds is measured from the acceptor or
 CC donor moiety or maybe obtained using a ratio between the two.

XX Sequence 238 AA;
 Query Match 98.6%; Score 1258; DB 18; Length 238;
 Best Local Similarity 98.7%; Pred. NO. 2e-123;
 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYGKILKFICTTGKLPVPMPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYGKILKFICTTGKLPVPMPTL 60
 QY 61 VTLSYGVQCFSRYPDHMKRHDFEFSKAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTLSYGVQCFSRYPDHMKRHDFEFSKAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRTELKGIQDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKVNFKIRHNIEDGGVQLAD 180
 DB 121 NRTELKGIQDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKVNFKIRHNIEDGGVQLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLGFVTAAGITHGMDELYK 238
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLGFVTAAGITHGMDELYK 238

RESULT 6
 AAW76371
 ID AAW76371 standard; Protein: 238 AA.
 XX AC AAW76371;
 XX DI 03-DEC-1998 (first entry)
 XX DE A. victoria green fluorescent protein.
 XX KW Green fluorescent protein; GFP; jellyfish; detection; mutant; activity;
 KW distribution.
 XX OS Aequorea victoria.
 XX PN JPI0234382-A.
 XX PD 08-SEP-1998.
 XX PF 27-FEB-1997; 97JP-0062370.
 XX PR 27-FEB-1997; 97JP-0062370.
 XX PA (JINA-) DINABEKKU KENYUSHO KK.
 XX PA WP: 1998-535036/46.
 XX DR N-PSDB; AAV61751.
 XX PT Green fluorescent protein - useful for determining promoter and
 PT targeting activity in a cell
 XX PS Disclosure; Page 6-7; 11pp; Japanese.

XX This sequence represents a wild-type green fluorescent protein, GFP,
 CC isolated from Aequorea victoria. This protein is used in the
 CC construction of a mutant protein which emits fluorescence 6 times
 CC higher than the conventional fluorescent proteins. The nucleic acid can
 CC be used for the recombinant production of the fluorescent protein and
 CC also for the determination of activity of a promoter to be tested,
 CC comprising steps of introducing the above vector to a host cell; and

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:06:07 ; Search time 41.333 Seconds

(without alignments)

1196.433 Million cell updates/sec

Title: US-09-967-301-3

Perfect score: 1276

Sequence: MSKGEELFTGVVPIVLVDG.....VLISFVTAAGIHGMDGLYK 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 57:580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 471540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP-archaea.*
- 2: SP-bacteria.*
- 3: SP-fungi.*
- 4: SP-human.*
- 5: SP-invertebrate.*
- 6: SP-mammal.*
- 7: SP-misc.*
- 8: SP-organelle.*
- 9: SP-phage.*
- 10: SP-plant.*
- 11: SP-toeplitz.*
- 12: SP-virus.*
- 13: SP-vertebrate.*
- 14: SP-unclassified.*
- 15: SP-virus.*
- 16: SP-bacteriophage.*
- 17: SP-archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1254	96.3	238	5 Q27903	Q27903 unidentified
2	1247	97.7	238	5 Q31225	Q31225 aquorea vi
3	1212	95.0	238	5 Q17105	Q17105 aquorea vi
4	1204	94.4	758	4 Q56J01	Q56J01 homo sapien
5	1197	93.8	238	5 Q17106	Q17106 aquorea vi
6	1105	86.6	238	5 Q8WTC6	Q8WTC6 aquorea ma
7	1101	86.3	238	5 Q8WP95	Q8WP95 aquorea ma
8	1097	86.0	238	5 Q8WTC4	Q8WTC4 aquorea ma
9	1095	85.8	238	5 Q8WTC0	Q8WTC0 aquorea ma
10	1094	85.7	238	5 Q8WTC9	Q8WTC9 aquorea ma
11	1094	85.7	238	5 Q8WTC8	Q8WTC8 aquorea ma
12	1092	85.6	238	5 Q8WTC7	Q8WTC7 aquorea ma
13	1090	85.4	238	5 Q8WTC5	Q8WTC5 aquorea ma
14	252.5	19.8	225	5 Q95UA7	Q95UA7 montastraea
15	250.5	19.6	236	5 Q8T600	Q8T600 dendronephr
16	250	19.6	225	5 Q963F5	Q963F5 montastraea

17	244.5	19.2	266	5 Q9U6Y3	Q9U6Y3 Clavularia
18	237.5	18.6	225	5 Q8T5F1	Q8T5F1 montastraea
19	211.5	16.6	234	5 Q8T5F2	Q8T5F2 montastraea
20	210.5	16.5	229	5 Q9U6Y5	Q9U6Y5 anemonia ma
21	210	16.5	222	5 Q95P04	Q95P04 gonioporta t
22	209	16.4	227	5 Q962P9	Q962P9 montastraea
23	207.5	16.3	225	5 Q9T6Y8	Q9T6Y8 discosoma s
24	205.5	15.1	232	5 Q9GP15	Q9GP15 anemonia su
25	205	16.1	227	5 Q95V10	Q95V10 montastraea
26	205	16.1	236	5 Q9BLY9	Q9BLY9 renilla muo
27	203.5	15.9	232	5 Q9U6Y7	Q9U6Y7 discosoma s
28	201.5	15.8	232	5 Q9GZ28	Q9GZ28 anemonia su
29	199.5	15.6	235	5 Q8T5F0	Q8T5F0 scolymia cu
30	197.5	15.5	225	5 Q8T6T9	Q8T6T9 heteractis
31	192	15.0	230	5 Q9GI37	Q9GI37 discosoma s
32	188.5	14.8	227	5 Q95W85	Q95W85 heteractis
33	188	14.7	228	5 Q9GP16	Q9GP16 anemonia su
34	184	14.4	233	5 Q963I9	Q963I9 renilla rec
35	182.5	14.3	238	5 Q9HLZ0	Q9HLZ0 pillosaurus
36	181.5	14.2	231	5 Q9U6Y5	Q9U6Y5 zoanthus sp
37	180.5	14.1	227	5 Q95W86	Q95W86 coelactis
38	174.5	13.7	227	5 Q95W11	Q95W11 coelactis
39	174.5	13.7	231	5 Q8T5E8	Q8T5E8 ridocea fl
40	172.5	13.5	229	5 Q8T5E7	Q8T5E7 coelactis
41	172	13.5	231	5 Q9U6Y4	Q9U6Y4 zoanthus sp
42	164.5	12.9	231	5 Q8T4U4	Q8T4U4 zoanthus sp
43	162.5	12.7	234	5 Q8C5F3	Q8C5F3 scolymia cu
44	153.5	12.0	231	5 Q8T6T8	Q8T6T8 discosoma s
45	143	11.2	231	5 Q8T5E9	Q8T5E9 ridocea fl

ALIGNMENTS

RESULT 1
Q27903 ID Q27903 PRELIMINARY: PRT: 236 AA.
AC Q27903;
CT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DI 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID=32644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9799032; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RC "Enhanced expression in tobacco of the gene encoding green fluorescent
RT protein by modification of its codon usage."
RI Plant Mol. Biol. 33:989-999(1997).
DR EMBL: X96418; CAA65278.1;
DR HSSP: P42212; LGP.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP.1.
DR PRINTS: PR01229; GFPDRESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 26884 MW: CA932047262AF2D3 CRC64:

Query Match: 98.3% Score 1254; DB 5; Length 238;
Best Local Similarity 98.3%; Pred. No. 3.2e-97;
Matches 234; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGCDATYGLTKLFCCTTGKLPVPMPTL 60
DB 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGCDATYGLTKLFCCTTGKLPVPMPTL 60
QY 61 VTLSYGVCFSKRYEDHMKRHDFKFSAMPEGYVQERTIFFKDDGNKTRAEVKEGDTLV 120
DB 61 VTLSYGVCFSKRYEDHMKRHDFKFSAMPEGYVQERTIFFKDDGNKTRAEVKEGDTLV 120

```

QY 121 NRIELKGIQKEDGNILGHKLEYNNSHVYIMADKCKNGIKYKLRHINISDGGVQLAD 180
DB 122 NRIELKGIQKEDGNILGHKLEYNNSHVYIMADKCKNGIKYKLRHINISDGGVQLAD 180
QY 181 HYQONTPIGDCPVLLPDNHYLSQSALSKDPNKRHMVLLGFTVAAGITHGMDELYK 238
DB 181 HYQONTPIGDCPVLLPDNHYLSQSALSKDPNKRHMVLLGFTVAAGITHGMDELYK 238

RESULT 2
Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125; 1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DR 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
CX NCBI_TaxID=6100;
RN [1]
RP MEDLINE=96305137; PubMed=8707053;
RA Carmack B.P., Valdivia R.H., Falkow S.;
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
[2]
SEQUENCE FROM N.A.
RA Carmack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.;
RA Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
RT expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73931; AB18957.1;
DR HSSP; P42212; iBFP.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 25940 MW; A2852280A9NEAF CRC64.

Query Match 97.74; Score 1247; DB 5; Length 238;
Best Local Similarity 97.59; Pred. No. 1, 2e-95;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPVILVELDGVNCHKFSVSGEGDGYGKILKFKICTGKLPVFWPTL 60
DB 1 MSKGEELFTGVVPVILVELDGVNCHKFSVSGEGDGYGKILKFKICTGKLPVFWPTL 60
QY 61 VTTLISYGVQCFSRYPDHMKRHDPEKFSAMPEGYVOERTIFFKDDGNYKTRAEVAFEDTLV 120
DB 61 VTTLISYGVQCFSRYPDHMKRHDPEKFSAMPEGYVOERTIFFKDDGNYKTRAEVAFEDTLV 120
QY 121 NRIELKGIQKEDGNILGHKLEYNNSHVYIMADKCKNGIKYKLRHINISDGGVQLAD 180
DB 121 NRIELKGIQKEDGNILGHKLEYNNSHVYIMADKCKNGIKYKLRHINISDGGVQLAD 180
QY 181 HYQONTPIGDCPVLLPDNHYLSQSALSKDPNKRHMVLLGFTVAAGITHGMDELYK 238
DB 181 HYQONTPIGDCPVLLPDNHYLSQSALSKDPNKRHMVLLGFTVAAGITHGMDELYK 238

RESULT 3
Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DR 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.

```

```

OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
CX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Katkios J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X63959; CAA58789.1;
DR HSSP; P42212; iBFP.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BL450E748E44 CRC64.

Query Match 95.04; Score 1212; DB 5; Length 238;
Best Local Similarity 94.18; Pred. No. 1e-93;
Matches 224; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPVILVELDGVNCHKFSVSGEGDGYGKILKFKICTGKLPVFWPTL 60
DB 1 MSKGEELFTGVVPVILVELDGVNCHKFSVSGEGDGYGKILKFKICTGKLPVFWPTL 60
QY 61 VTTLISYGVQCFSRYPDHMKRHDPEKFSAMPEGYVOERTIFFKDDGNYKTRAEVAFEDTLV 120
DB 61 VTTLISYGVQCFSRYPDHMKRHDPEKFSAMPEGYVOERTIFFKDDGNYKTRAEVAFEDTLV 120
QY 121 NRIELKGIQKEDGNILGHKLEYNNSHVYIMADKCKNGIKYKLRHINISDGGVQLAD 180
DB 121 NRIELKGIQKEDGNILGHKLEYNNSHVYIMADKCKNGIKYKLRHINISDGGVQLAD 180
QY 181 HYQONTPIGDCPVLLPDNHYLSQSALSKDPNKRHMVLLGFTVAAGITHGMDELYK 238
DB 181 HYQONTPIGDCPVLLPDNHYLSQSALSKDPNKRHMVLLGFTVAAGITHGMDELYK 238

RESULT 4
Q96J01 PRELIMINARY; PRT; 756 AA.
AC Q96J01;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DR 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 25, Last annotation update)
DE Raichu404X.
GN Raichu404X.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN.
KK MEDLINE=21322811; PubMed=11429608;
RA Kocizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
RA Miyawaki A., Matsuda M.;
RT "Spacio-temporal Images of Growth Factor-induced Activation of Ras and
RT Rap1.";
RI Nature 411:1065-1068(2001);
DR EMBL; AB051845; BAB01868.1;
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnstrimg.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01353; GFP; 2.
DR Pfam; PF00071; ras; 1.
DR Pfam; PF02196; RBD; 1.
DR PRODOM; PD013756; Green_fl_protein; 2.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW GTP-binding.

```

Q8WTC6;
 C1-MAR-2002 (TrEMBLrel. 20. Created)
 C1-MAR-2002 (TrEMBLrel. 20. Last sequence update)
 01-JUN-2002 (TrEMBLrel. 21. Last annotation update)
 Green fluorescent protein.
 GFP.
 CS Aequorea macrodactyla.
 OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OC NCBI_TaxID=147615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GFPXMI9UV.
 RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Gao B.O., Chen M.,
 Li S.J., Xia N.S.;
 RT "Colorful mutants of green fluorescent protein from Aequorea
 macrodactyla.";
 RC Submitted (0CT-2001) to the EMBL/GenBank/DBJ databases.
 RC ENBL; AF435431; AAL35916.1;
 DR InterPro: IPR000786; Green_fl_protein.
 DK Pfam: PF01353; GFP; 1
 DK ProDom: PD013756; Green_fl_protein; 1
 SQ SEQUENCE 238 AA: 270.5 MW; 6B9FD75E69526993 CRC64;
 Query Match 86.3%; Score 1105; DP 5; Length 238;
 Best Local Similarity 83.2%; Pred. No. 9; 5e-85;
 Matches 139; Conservative 19; Mismatches 20; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPILVELDGDVNHKFSVSGEGDAATYKRIILKFICTISKLVPKPTL 60
 DL 1 MSKGEELFTGVPIILVELDGDVNHKFSVSGEGDAATYKRIILKFICTISKLVPKPTL 60
 QY 61 VTLSYGVQCISRYPDHMKRHDFFKSAMPQGYVQERLIPFDQGNKYKRAEVKPSGITLV 120
 DL 61 VTLSYGLQCARYPERHKKNDFFKSAMPQGYVQERLIPFDQGNKYKRAEVKPSGITLV 120
 QY 121 NRIELKIDGFEDGNIIGHKLEYNYNSHNYIMADKKNGIKYVKIRINIEGGVQLAD 180
 DL 121 NRIELKIDGFEDGNIIGHKLEYNYNSHNYIMADKKNGIKYVKIRINIEGGVQLAD 180
 QY 181 HYQNTPTGDDGVPILPNHYSTQSALSCKPNCKRDMVLLGFTVAAGITHCMDELYK 238
 DL 181 HYQNTPTGDDGVPILPNHYLSTQALSCKPNCKRDMVLLGFTVAAGITHCMDELYK 238
 RESULT 7
 Q8WPF5 PRELIMINARY; PRT: 238 AA.
 LU Q8WPF5
 AC Q8WPF5;
 DT C1-MAR-2002 (TrEMBLrel. 20. Created)
 DT C1-MAR-2002 (TrEMBLrel. 20. Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21. Last annotation update)
 DE Green fluorescent protein.
 DE GFPX.
 CS Aequorea macrodactyla.
 OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OC NCBI_TaxID=147615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GFPXMI9UV.
 RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Gao B.O., Chen M.,
 Li S.J., Xia N.S.;
 RT "Colorful mutants of green fluorescent protein from Aequorea
 macrodactyla.";
 RC Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RC ENBL; AY013824; AAK02052.1;
 DR InterPro: IPR000786; Green_fl_protein.
 DK Pfam: PF01353; GFP; 1
 DK ProDom: PD013756; Green_fl_protein; 1
 SQ SEQUENCE 238 AA: 270.9 MW; 8185D0E5E529012A CRC64;
 Query Match 86.3%; Score 1101; DP 5; Length 238;
 Best Local Similarity 83.2%; Pred. No. 2e-84;

```

Matches 196; Conservative 13; Mismatches 21; Gaps 0;
QY 1 MSKGEELFTGVVPLVLELGDGVNKHKFSVSGEGSDATYCKLTKEICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELGDGVNKHKFSVSGEGSDATYCKLTKEICTTGKLPVWPPTL 60
QY 61 VTTLSYGVOCSRYPPDHKKRHDFFKSNPESYVQERTIFFKDDGNYKTRAEVKFEGDILV 120
DB 61 VTTLSYGVOCSRYPPDHKKRHDFFKSNPESYVQERTIFFKDDGNYKTRAEVKFEGDILV 120
QY 121 NRLEKGDIDFKEDGNILGHKLEYNNSHNYIMADKKNGKYNFKIRNIEHGQVQLAD 180
DB 121 NRLEKGDIDFKEDGNILGHKLEYNNSHNYIMADKKNGKYNFKIRNIEHGQVQLAD 180
QY 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
RESULT 6
Q8WTC4 PRELIMINARY: PRT: 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=SHG24;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435433; AAL330.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 26956 MW: 75521EAF0CB5A7AA CAG64;
Query Match 86.0%; Score 1097; DB 5; Length 238;
Best Local Similarity 84.0%; Pred. No. 4.6e-84;
Matches 200; Conservative 16; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPLVLELGDGVNKHKFSVSGEGSDATYCKLTKEICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELGDGVNKHKFSVSGEGSDATYCKLTKEICTTGKLPVWPPTL 60
QY 61 VTTLSYGVOCSRYPPDHKKRHDFFKSNPESYVQERTIFFKDDGNYKTRAEVKFEGDILV 120
DB 61 VTTLSYGVOCSRYPPDHKKRHDFFKSNPESYVQERTIFFKDDGNYKTRAEVKFEGDILV 120
QY 121 NRLEKGDIDFKEDGNILGHKLEYNNSHNYIMADKKNGKYNFKIRNIEHGQVQLAD 180
DB 121 NRLEKGDIDFKEDGNILGHKLEYNNSHNYIMADKKNGKYNFKIRNIEHGQVQLAD 180
QY 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
RESULT 9
Q8WTC9 PRELIMINARY: PRT: 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX162;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435428; AAL339.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27031 MW: 5F80A19C19DC584D CRC64;
Query Match 85.7%; Score 1094; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;

```

```

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435427; AAL339.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 26957 MW: 5F80A192173CH84D CRC64;
Query Match 85.8%; Score 1055; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 6.5e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPLVLELGDGVNKHKFSVSGEGSDATYCKLTKEICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELGDGVNKHKFSVSGEGSDATYCKLTKEICTTGKLPVWPPTL 60
QY 61 VTTLSYGVOCSRYPPDHKKRHDFFKSNPESYVQERTIFFKDDGNYKTRAEVKFEGDILV 120
DB 61 VTTLSYGVOCSRYPPDHKKRHDFFKSNPESYVQERTIFFKDDGNYKTRAEVKFEGDILV 120
QY 121 NRLEKGDIDFKEDGNILGHKLEYNNSHNYIMADKKNGKYNFKIRNIEHGQVQLAD 180
DB 121 NRLEKGDIDFKEDGNILGHKLEYNNSHNYIMADKKNGKYNFKIRNIEHGQVQLAD 180
QY 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
RESULT 10
Q8WTC9 PRELIMINARY: PRT: 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX162;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435428; AAL339.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27031 MW: 5F80A19C19DC584D CRC64;
Query Match 85.7%; Score 1094; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;

```

```

QY 1 MSKGEELFTGVVPIIVELDGVNCHKFSVSGEGDATYGLKTLAKTCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIIVELDGVNCHKFSVSGEGDATYGLKTLAKTCTTCKLPVWPPTL 60
QY 61 VTTLSYGVOCSRYPDHMKRHDFFKSAPEGYVQERTIFFKDGSKYKTRAEVKEFEDTIV 120
DB 61 VTTLSYGVOCSRYPDHMKRHDFFKSAPEGYVQERTIFFKDGSKYKTRAEVKEFEDTIV 120
QY 121 NRIELKGDFKEDGNILGHKLEYNHNHYIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
DB 121 NRIELKGDFKEDGNILGHKLEYNHNHYIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
QY 181 HYQONTPIGGPVLPPDNHYSIQSALSKDPNEKSDHNVLLGFTVTAAGITGMDLEYK 238
DB 181 HYQONTPIGGPVLPPDNHYSIQSALSKDPNEKSDHNVLLGFTVTAAGITGMDLEYK 238

RESULT 11
Q8WTC8 PRELIMINARY: PRT: 238 AA.
AC Q8WTC8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435430; AAL33915.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F60A18FAE7C84D CAC64;

Query Match 85.7%; Score 1094; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F60A18FAE7C84D CAC64;

QY 1 MSKGEELFTGVVPIIVELDGVNCHKFSVSGEGDATYGLKTLAKTCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIIVELDGVNCHKFSVSGEGDATYGLKTLAKTCTTCKLPVWPPTL 60
QY 61 VTTLSYGVOCSRYPDHMKRHDFFKSAPEGYVQERTIFFKDGSKYKTRAEVKEFEDTIV 120
DB 61 VTTLSYGVOCSRYPDHMKRHDFFKSAPEGYVQERTIFFKDGSKYKTRAEVKEFEDTIV 120
QY 121 NRIELKGDFKEDGNILGHKLEYNHNHYIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
DB 121 NRIELKGDFKEDGNILGHKLEYNHNHYIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
QY 181 HYQONTPIGGPVLPPDNHYSIQSALSKDPNEKSDHNVLLGFTVTAAGITGMDLEYK 238
DB 181 HYQONTPIGGPVLPPDNHYSIQSALSKDPNEKSDHNVLLGFTVTAAGITGMDLEYK 238

RESULT 12
Q8WTC7 PRELIMINARY: PRT: 238 AA.
AC Q8WTC7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F60A18FAE7C84D CAC64;

Query Match 85.7%; Score 1094; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F60A18FAE7C84D CAC64;

```

```

GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435430; AAL33915.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F5BA2982264C0E CRC64;

Query Match 85.6%; Score 1092; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 1.2e-83;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIIVELDGVNCHKFSVSGEGDATYGLKTLAKTCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIIVELDGVNCHKFSVSGEGDATYGLKTLAKTCTTCKLPVWPPTL 60
QY 61 VTTLSYGVOCSRYPDHMKRHDFFKSAPEGYVQERTIFFKDGSKYKTRAEVKEFEDTIV 120
DB 61 VTTLSYGVOCSRYPDHMKRHDFFKSAPEGYVQERTIFFKDGSKYKTRAEVKEFEDTIV 120
QY 121 NRIELKGDFKEDGNILGHKLEYNHNHYIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
DB 121 NRIELKGDFKEDGNILGHKLEYNHNHYIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
QY 181 HYQONTPIGGPVLPPDNHYSIQSALSKDPNEKSDHNVLLGFTVTAAGITGMDLEYK 238
DB 181 HYQONTPIGGPVLPPDNHYSIQSALSKDPNEKSDHNVLLGFTVTAAGITGMDLEYK 238

RESULT 13
Q8WTC5 PRELIMINARY: PRT: 238 AA.
AC Q8WTC5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Orange fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435432; AAL33917.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534573A CRC64;

Query Match 85.4%; Score 1090; DB 5; Length 238;
Best Local Similarity 83.6%; Pred. No. 1.7e-83;
Matches 199; Conservative 16; Mismatches 23; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIIVELDGVNCHKFSVSGEGDATYGLKTLAKTCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIIVELDGVNCHKFSVSGEGDATYGLKTLAKTCTTCKLPVWPPTL 60

```

Do 1 MSKGEELFTGVVPLVFLDGVNKHKFSVSGSGANDYKLEIKFICTL3KLPVWPTL 60
 Qy 61 VTTSYSGVGGESRYVDHMKHDFKFSAMPESYVQERTIFFKQDGNKYKTRAEVKEPNDITV 120
 Db 61 VTTSYSGVGGESRYVDHMKHDFKFSAMPESYVQERTIFFKQDGNKYKTRAEVKEPNDITV 120
 Qy 121 NRILEKGLDFKEDGNILGHLEKYNHSHNYIMADQKNGIKYNFKLRINIEDGGVOLAD 180
 Db 121 NRILEKGLDFKEDGNILGHLEKYNHSHNYIMADQKNGIKYNFKLRINIEDGGVOLAD 180
 Qy 181 HYQNTPTGDPVLLPNDHNLSTQSALSQDNKRDHMLVGSVTAAGITHGMDLEYK 238
 Db 181 HYQNTPTGDPVLLPNDHNLSTQSALSQDNKRDHMLVGSVTAAGITHGMDLEYK 238
 RESULT 14
 Q950A7 PRELIMINARY: PRT: 225 AA.
 AC Q950A7
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Cyan fluorescent protein (Pradment).
 OS Montastraea cavernosa (great star coral).
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Scleractinia;
 OC Faviidae; Faviidae; Montastraea.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Falkowski P.G., Sun Y.
 RT "Montastraea cavernosa fluorescent protein."
 RC Submitted (Sep-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY056460; AAL17505.1;
 DR InterPro: IPR00786; Green_fl_protein.
 DR Pfam: PF01353; GFP. 1.
 DR ProDom: PD01756; Green_fl_protein; 1.
 FT NON_TER 225
 FT NOK_TER 225
 SQ SEQUENCE 225 AA: 25775 MW: 5216271400.9524 DPK64;

Query Match 19.6%; Score 252.5; DB 5; Length 225;
 Best Local Similarity 28.3%; Pred. No. 20:19;
 Matches 67; Conservative 48; Mismatches 19; Indels 41; Gaps 9;
 QY 11 VVPIILVELDGVNKHKFSVSGEGDATYKTLKFIKTTGKLPVWPTLVTLSYGVQ 69
 Db 7 VMKIKRMGIVNGHKNITGEGCKPEGTHTTLKVKSGRIPEAYDICTAFQVGNR 66
 QY 70 QPSRYPDHMKHDFKFSAMPESYVQERTIFFKQDGNKYKTRAEVKEPNDITVNP 124
 Db 67 VFTKYPKDIP--DYFKQSPPEGYSWERSMTFEDGVGVTVISDKLRDCTFEY 124
 QY 130 FKEDGNILGHK--LEYNYSNHNVIYIMADQKNGIKYNFKLRINIEDGGVOLAD 180
 Db 125 FSSGVPVNGKTKLWEPSTENKVV-----RDGVLLGVSKTLLIEDG-----KH 175
 QY 189 GGSGP---VLLDHNHVLSTQSALSQDNKRDHMLVGSVTAAGITHGMD--- 238
 Db 176 TYSARKGVLPVPEYFV-----DHRKEFI-----LSHKKQYNTVEVYE 211

RESULT 15
 Q8T600 PRELIMINARY: PRT: 235 AA.
 ID Q8T600
 AC Q8T600;
 DT 01-JUN-2002 (TRENBLrel. 21, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Green fluorescent protein.
 OS Dendronephthya sp. SGAJ-2002.
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
 OC Nephtheidae; Dendronephthya.
 OX NCBI_TaxID=191210;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21927629; PubMed=11929995;
 RA Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
 RA Lukyanov K.A., Lukyanov S.A., Matz M.V.;
 RT "Diversity and evolution of the green fluorescent protein family."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
 DR EMBL: AF420591; AAM10625.1;
 SQ SEQUENCE 235 AA: 26840 MW: 5217070799334490 CRC64;
 Query Match 19.6%; Score 250.5; DB 5; Length 235;
 Best Local Similarity 31.8%; Pred. No. 3:2e-13;
 Matches 64; Conservative 39; Mismatches 85; Indels 13; Gaps 7;
 QY 14 ILVE-DGDVNGHKFSVSGEGDATYKTLKFIKTTG-KLPVWPTLVTLSYGVQCES 72
 Db 10 VKVHMEGNVNGHAFVIEGSGKRPYEGTCTLNLTVKSGAPLPFSYD:ITTA:HYGRVFT 69
 QY 73 RYPDHMKHDFKFSAMPESYVQERTIFFKQDGNKYKTRAEVKEPNDITVNP 132
 Db 70 EYPADIT--DYFKQSPPEGYSWERSMTFEDGVGVTVISDKLRDCTFEY 127
 QY 133 DGNILGHK--LEYNYSNHNVIYIMADQKNGIKV-NFKIRHNIEDGGVOLADHYQNTPI 190
 Db 128 KQVNMCKTKLWEPSTENKVV-----RDGVLLGVSKTLLIEDG-----KH 175
 QY 191 GPVLLPDNHYLSTQ-SA:SKD 210
 Db 181 KVVQIPDVHFVDRRIEILSND 201

Search completed: July 16, 2003. 16:13:29
 Job time : 42.333 secs

Genore version 5.1.6
Copyright (c) 1993 - 2003 CompuLink Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:11:38 : Search time 29 Seconds

(Without alignments)

974,649 Million cells - 463,435/sec

Title: US-09-967-301-4

Perfect score: 1276

Sequence: 1 MSKGEELFTGVVPLVELDGLG.....VLLRVVANGIHHGMELEK 238

Scoring table: RLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759775 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	1262	98.9	238	9	US-09-967-301-2	Sequence 2, Appl
2	1262	98.9	238	10	US-09-884-681-2	Sequence 2, Appl
3	1262	98.9	238	12	US-10-024-636-2	Sequence 2, Appl
4	1262	98.9	1070	9	US-10-001-486-2	Sequence 2, Appl
5	1262	98.9	1099	9	US-10-253-864-6	Sequence 4, Appl
6	1262	98.9	1147	9	US-10-253-864-1	Sequence 1, Appl
7	1262	98.3	1452	12	US-10-050-673-2	Sequence 2, Appl
8	1259	98.7	238	9	US-09-899-954B-2	Sequence 2, Appl
9	1258	98.6	238	9	US-09-856-518-2	Sequence 2, Appl
10	1258	98.6	238	9	US-09-900-345A-125	Sequence 125, App
11	1258	98.6	238	9	US-10-121-451-10	Sequence 10, App
12	1258	98.6	238	10	US-09-920-922-4	Sequence 6, Appl
13	1258	98.6	238	10	US-09-852-900-1	Sequence 4, Appl
14	1258	98.4	244	9	US-09-899-954B-8	Sequence 1, Appl
15	1256	98.4	244	9	US-09-899-954B-10	Sequence 10, Appl
16	1256	98.4	244	9	US-09-899-954B-12	Sequence 12, Appl
17	1256	98.4	244	9	US-09-899-954B-14	Sequence 14, Appl
18	1256	98.4	244	9	US-09-899-954B-16	Sequence 16, Appl
19	1256	98.4	244	9	US-09-999-954B-16	Sequence 16, Appl

20	1256	98.4	244	9	US-09-999-954B-18	Sequence 18, Appl
21	1256	98.4	238	9	US-09-889-025A-2	Sequence 2, Appl
22	1256	98.4	243	9	US-09-900-345A-60	Sequence 60, Appl
23	1256	98.4	243	9	US-09-900-345A-62	Sequence 62, Appl
24	1256	98.4	243	9	US-09-900-345A-64	Sequence 64, Appl
25	1256	98.4	243	9	US-09-900-345A-66	Sequence 66, Appl
26	1256	98.4	243	9	US-09-900-345A-68	Sequence 68, Appl
27	1256	98.4	243	9	US-09-900-345A-70	Sequence 70, Appl
28	1254	98.3	237	9	US-09-899-954B-36	Sequence 36, Appl
29	1254	98.3	243	9	US-09-900-345A-56	Sequence 56, Appl
30	1254	98.3	243	9	US-09-900-345A-58	Sequence 58, Appl
31	1254	98.3	243	9	US-09-900-345A-114	Sequence 114, App
32	1254	98.3	243	9	US-09-900-345A-116	Sequence 116, App
33	1254	98.3	243	9	US-09-900-345A-118	Sequence 118, App
34	1254	98.3	243	9	US-09-900-345A-120	Sequence 120, App
35	1254	98.3	243	9	US-09-899-954B-4	Sequence 4, Appl
36	1254	98.3	244	9	US-09-899-954B-6	Sequence 6, Appl
37	1254	98.3	244	9	US-09-899-954B-20	Sequence 20, Appl
38	1254	98.3	244	9	US-09-900-345A-2	Sequence 2, Appl
39	1254	98.3	244	9	US-09-900-345A-4	Sequence 4, Appl
40	1253	98.2	243	9	US-09-900-345A-6	Sequence 6, Appl
41	1253	98.2	243	9	US-09-900-345A-8	Sequence 8, Appl
42	1253	98.2	243	9	US-09-900-345A-10	Sequence 10, Appl
43	1253	98.2	243	9	US-09-900-345A-12	Sequence 12, Appl
44	1253	98.2	243	9	US-09-900-345A-12	Sequence 12, Appl
45	1253	98.2	243	9	US-09-900-345A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-10-057-505-2
: Sequence 2, Application US/10057505
: Patent No. US20020164674A1
: GENERAL INFORMATION:
: APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
: APPLICANT: AURORA BIOSCIENCES CORPORATION
: APPLICANT: TSIEN, Roger
: APPLICANT: HEIM, Roger
: APPLICANT: CURITT, Andrew
: TITLE OF INVENTION: FANCDEN FLUORESCENT PROTEIN CONSTRUCTS
: FILE REFERENCE: REG1260-3
: CURRENT APPLICATION NUMBER: US/10/957,505
: PRIOR FILING DATE: 2002-01-25
: PRIOR APPLICATION NUMBER: US 08/792,553
: PRIOR FILING DATE: 1997-01-31
: PRIOR APPLICATION NUMBER: US 09/396,003
: PRIOR FILING DATE: 1999-09-13
: NUMBER OF SEQ ID NOS: 31
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 238
: TYPE: PPT
: ORGANISM: Acetobacter victoria
US-10-057-505-2

Query Match 98.9% Score 1262: DB 9: Length 238:

Best Local Similarity 95.7%: Pred. No. 2,7e-115:

Mismatches 1: Mismatches 2: Indels 0: Gaps 0:

Matches 235: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

QY 1 MSKGEELFTGVVPLVELDGLG.....VLLRVVANGIHHGMELEK 238

DB 1 MSKGEELFTGVVPLVELDGLG.....VLLRVVANGIHHGMELEK 238

QY 6: VTLTYGVQCFSSRYPDHMKRDEFEKSAMPEGYVQERTIFPKDGNKYKTRAEKFFEDTLY 120

DB 6: VTLTYGVQCFSSRYPDHMKRDEFEKSAMPEGYVQERTIFPKDGNKYKTRAEKFFEDTLY 120

QY 121 NRTELKGDIFKEFGNLSGHKLEYNVNSHNHYINADKKNIGIKVNFIRHNIIEGGVGLAD 180

DB 121 NRTELKGDIFKEFGNLSGHKLEYNVNSHNHYINADKKNIGIKVNFIRHNIIEGGVGLAD 180

QY 181 HYQONTPIGDPVLLPNDNYHSTOSALSCKDPNEKEHNVLLFEVTAAGITHGMDELYK 238
|||||.....
DB 181 HYQONTPIGDPVLLPNDNYHSTOSALSCKDPNEKEHNVLLFEVTAAGITHGMDELYK 238

RESULT 2

US-09-884-681-2

Sequence 2, Application US/09884681

Patent No. US20020061546A1

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

Cubitt, Andrew B.

TITLE OF INVENTION: Assays for Protein Kinases Using

Fluorescent Protein Substrates

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/884,681

FILING DATE: 19-Jun-2001

CLASSIFICATION: <Unknown>

Prior APPLICATION DATA:

APPLICATION NUMBER: 08/675,865

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John S.

REGISTRATION NUMBER: 32,944

REFERENCE/DOCKET NUMBER: 023072-06-000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-884-681-2

Query Match

Best Local Similarity 98.9% Score 1262 DB 10 Length 238

Matches 235: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

QY 1 MSKGEELFTGVPLTLLVDGVNGHKFSVSGEGGDAVYKLTIKFKICTIGKLPVWPPL 60
|||.....
DB 1 MSKGEELFTGVPLTLLVDGVNGHKFSVSGEGGDAVYKLTIKFKICTIGKLPVWPPL 60
|||.....
QY 61 VTILYGVQCFSRYPDHMKRDEDFKSAMPEGYVQERTIFFKDDGNYKTRAEYKFEGLTIV 120
|||.....
DB 61 VTILYGVQCFSRYPDHMKRDEDFKSAMPEGYVQERTIFFKDDGNYKTRAEYKFEGLTIV 120
|||.....
QY 121 NRIELKIDPKEDNTLGHKLEYNYNHNYIMADKQKGIKYNFKIRNIEGSGVOLAD 180
|||.....
DB 121 NRIELKIDPKEDNTLGHKLEYNYNHNYIMADKQKGIKYNFKIRNIEGSGVOLAD 180
|||.....
QY 181 HYQONTPIGDPVLLPNDNYHSTOSALSCKDPNEKEHNVLLFEVTAAGITHGMDELYK 238
|||.....
DB 181 HYQONTPIGDPVLLPNDNYHSTOSALSCKDPNEKEHNVLLFEVTAAGITHGMDELYK 238

RESULT 3

US-10-024-686-2

Sequence 2, Application US/10001496B

Publication No. US20030082642A2

Sequence 2, Application US/10024686

Patent No. US20020123113A2

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

Heim, Roger

TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/024,686

FILING DATE: 17-Dec-2001

Prior APPLICATION DATA:

APPLICATION NUMBER: 09/557,935

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/727,452

FILING DATE: 10-OCT-1996

APPLICATION NUMBER: US95/14652

FILING DATE: 13-NOV-1995

APPLICATION NUMBER: 08/337,915

FILING DATE: 10-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.

REGISTRATION NUMBER: 35,147

REFERENCE/DOCKET NUMBER: 02257/032022

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/678-5070

TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-024-686-2

Query Match

Best Local Similarity 98.9% Score 1262 DB 12 Length 238

Matches 235: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

QY 1 MSKGEELFTGVPLTLLVDGVNGHKFSVSGEGGDAVYKLTIKFKICTIGKLPVWPPL 60
|||.....
DB 1 MSKGEELFTGVPLTLLVDGVNGHKFSVSGEGGDAVYKLTIKFKICTIGKLPVWPPL 60
|||.....
QY 61 VTILYGVQCFSRYPDHMKRDEDFKSAMPEGYVQERTIFFKDDGNYKTRAEYKFEGLTIV 120
|||.....
DB 61 VTILYGVQCFSRYPDHMKRDEDFKSAMPEGYVQERTIFFKDDGNYKTRAEYKFEGLTIV 120
|||.....
QY 121 NRIELKIDPKEDNTLGHKLEYNYNHNYIMADKQKGIKYNFKIRNIEGSGVOLAD 180
|||.....
DB 121 NRIELKIDPKEDNTLGHKLEYNYNHNYIMADKQKGIKYNFKIRNIEGSGVOLAD 180
|||.....
QY 181 HYQONTPIGDPVLLPNDNYHSTOSALSCKDPNEKEHNVLLFEVTAAGITHGMDELYK 238
|||.....
DB 181 HYQONTPIGDPVLLPNDNYHSTOSALSCKDPNEKEHNVLLFEVTAAGITHGMDELYK 238

RESULT 4

US-10-001-486B-2

Sequence 2, Application US/10001496B

Publication No. US20030082642A2

GENERAL INFORMATION:
APPLICANT: The Government of the United States of America
as represented by the Secretary
Department of Health and Human Services
Washington, D.C.
Htun Ph.D., Han
Hager Ph.D., Gordon L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MANIPULATING
DNA BINDING MOLECULES IN LIVING CELLS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.0
CURRENT APPLICATION NUMBER: US/16/001,486B
FILING DATE: 15-NOV-2003/052642A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,373
FILING DATE: 08 Dec 1995
ATTORNEY/AGENT INFORMATION:
NAME: Selby, Elizabeth
REGISTRATION NUMBER: 38298
REFERENCE/DOCKET NUMBER: 14014.0183
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-688-0770
TELEFAX: 404-688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-16-001-486B-2
Query Match 98.9% Score 1262; DB 9; Length 1079;
Best Local Similarity 98.7% Pred. No. 2,1e-114;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
61 VTITTYGVQCFSSRYPDHMKRHDFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
61 VTITTYGVQCFSSRYPDHMKRHDFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
121 NRTELKGDIFKEDGNILGHKLEYNNSHNYVIMAKQKNGIKVNFIRHNLEDGSQLAD 160
121 NRTELKGDIFKEDGNILGHKLEYNNSHNYVIMAKQKNGIKVNFIRHNLEDGSQLAD 180
181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVVLEFVTAAGITHGMDELYK 238
181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVVLEFVTAAGITHGMDELYK 238
RESULT 5
US-10-259-864-1
Sequence 1, Application US/10/259864
Publication No. US20030077645A1
GENERAL INFORMATION:
APPLICANT: Hager, Gordon L.
TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfa
TITLE OF INVENTION: Receptor Ligands, and Methods and Kits For Detecting and Char
FILE REFERENCE: 24799
CURRENT APPLICATION NUMBER: US/10/259,864
PRIOR FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: US 60/325,178
PRIOR FILING DATE: 2001-09-28
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 1099
TYPE: PRT
ORGANISM: Rat/Human Chimera
US-10-259-864-4
Query Match 98.9% Score 1262; DB 9; Length 1099;
Best Local Similarity 98.7% Pred. No. 2,1e-114;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
61 VTITTYGVQCFSSRYPDHMKRHDFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
61 VTITTYGVQCFSSRYPDHMKRHDFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
121 NRTELKGDIFKEDGNILGHKLEYNNSHNYVIMAKQKNGIKVNFIRHNLEDGSQLAD 160
121 NRTELKGDIFKEDGNILGHKLEYNNSHNYVIMAKQKNGIKVNFIRHNLEDGSQLAD 180
181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVVLEFVTAAGITHGMDELYK 238
181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVVLEFVTAAGITHGMDELYK 238
RESULT 5
US-10-259-864-4
Sequence 4, Application US/10/259864
Publication No. US20030077645A1
GENERAL INFORMATION:
APPLICANT: Hager, Gordon L.
TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfamili

1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
61 VTITTYGVQCFSSRYPDHMKRHDFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
61 VTITTYGVQCFSSRYPDHMKRHDFKFSAMPEGVVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
121 NRTELKGDIFKEDGNILGHKLEYNNSHNYVIMAKQKNGIKVNFIRHNLEDGSQLAD 160
121 NRTELKGDIFKEDGNILGHKLEYNNSHNYVIMAKQKNGIKVNFIRHNLEDGSQLAD 180
181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVVLEFVTAAGITHGMDELYK 238
181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHVVLEFVTAAGITHGMDELYK 238
RESULT 5
US-10-259-864-1
Sequence 1, Application US/10/259864
Publication No. US20030077645A1
GENERAL INFORMATION:
APPLICANT: Hager, Gordon L.
TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfa
TITLE OF INVENTION: Receptor Ligands, and Methods and Kits For Detecting and Char
FILE REFERENCE: 24799
CURRENT APPLICATION NUMBER: US/10/259,864
PRIOR FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: US 60/325,178
PRIOR FILING DATE: 2001-09-28
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1147
TYPE: PPT
ORGANISM: Rat/Human Chimera
FEATURE:
NAME/KEY: mat_peptide
LOCATION: (1)...(1)
OTHER INFORMATION: Chimeric Protein
US-10-259-864-1
Query Match 98.9% Score 1262; DB 9; Length 1147;
Best Local Similarity 98.7% Pred. No. 2,2e-114;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

CY 121 NRIELKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKEHNTEDGSGVLAD 180
DB 121 NRIELKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKEHNTEDGSGVLAD 180
CY 181 HYQONTPIGSGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITTHGMDELYK 238
DB 181 HYQONTPIGSGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITTHGMDELYK 238

RESULT 7
US-10-050-673-2
Sequence 2: Application US/10050673
Patent No. US2002015133A1
GENERAL INFORMATION:
APPLICANT: David M. Knipe
APPLICANT: Elizabeth McNamee
TITLE OF INVENTION: Replication-Competent Virus Expressing A
FILE REFERENCE: H098-05
CURRENT APPLICATION NUMBER: US/10-050-673
CURRENT FILING DATE: 2002-05-16
PRIOR APPLICATION NUMBER: US/99-127-227
PRIOR FILING DATE: 1998-07-31
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1452
TYPE: PRT
ORGANISM: herpesvirus
US-10-050-673-2

Query Match 98.9%; Score 1262; DB 12; Length 1452;
Best Local Similarity 98.7%; Pred. No. 36-114;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPIIVLVDGVDNGHGFSSVSGEGDATYGKLTIKFKICTTCKLVPWPTL 60
DB 1215 MSKGEELFTGVVPIIVLVDGVDNGHGFSSVSGEGDATYGKLTIKFKICTTCKLVPWPTL 1274
CY 61 VTITLVGVQCFSPYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNKYKTRAEVKFEGDTLV 126
DB 1275 VTITLVGVQCFSPYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNKYKTRAEVKFEGDTLV 1351
CY 121 NRIELKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKEHNTEDGSGVLAD 180
DB 1335 NRIELKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKEHNTEDGSGVLAD 1394
CY 181 HYQONTPIGSGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITTHGMDELYK 238
DB 1395 HYQONTPIGSGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITTHGMDELYK 1452

RESULT 8
US-09-999-954B-2
Sequence 2: Application US/9989954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Rotella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein in
FILE REFERENCE: 10338-50S
CURRENT APPLICATION NUMBER: US/99-999,954B
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU HP8677
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: PCT/AU00/0007
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 38
```

```

SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 238
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: GFP control construct
US-09-899-954B-2

Query Match 98.7%; Score 1259; DB 9; Length 238;
Best Local Similarity 96.3%; Pred. No. 53e-115;
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPIIVLVDGVDNGHGFSSVSGEGDATYGKLTIKFKICTTCKLVPWPTL 60
DB 1 MSKGEELFTGVVPIIVLVDGVDNGHGFSSVSGEGDATYGKLTIKFKICTTCKLVPWPTL 60
CY 61 VTITLVGVQCFSPYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNKYKTRAEVKFEGDTLV 120
DB 61 VTITLVGVQCFSPYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNKYKTRAEVKFEGDTLV 120
CY 121 NRIELKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKEHNTEDGSGVLAD 180
DB 121 NRIELKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKEHNTEDGSGVLAD 180
CY 181 HYQONTPIGSGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITTHGMDELYK 238
DB 181 HYQONTPIGSGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITTHGMDELYK 238

RESULT 9
US-09-866-538-2
Sequence 2: Application US/09866538
Publication No. US2003003268A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: SIEN, Roger
TITLE OF INVENTION: NON-GLYCERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530-2
CURRENT APPLICATION NUMBER: US/09-866,538
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-866-538-2

Query Match 98.6%; Score 1258; DB 9; Length 238;
Best Local Similarity 96.3%; Pred. No. 6.6e-115;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPIIVLVDGVDNGHGFSSVSGEGDATYGKLTIKFKICTTCKLVPWPTL 60
DB 1 MSKGEELFTGVVPIIVLVDGVDNGHGFSSVSGEGDATYGKLTIKFKICTTCKLVPWPTL 60
CY 61 VTITLVGVQCFSPYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNKYKTRAEVKFEGDTLV 120
DB 61 VTITLVGVQCFSPYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNKYKTRAEVKFEGDTLV 120
CY 121 NRIELKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKEHNTEDGSGVLAD 180
DB 121 NRIELKGIIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKEHNTEDGSGVLAD 180
CY 181 HYQONTPIGSGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITTHGMDELYK 238
DB 181 HYQONTPIGSGPVLLPDNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITTHGMDELYK 238

RESULT 10
US-09-960-345A-125
```

Sequence 125, Application US/05950345A

Publication No. US20030031369A1

GENERAL INFORMATION:

APPLICANT: Frazer, Ian Hector

APPLICANT: Zhou, Jian

TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL

TITLE OF INVENTION: EFFICIENCY OF A CODON

FILE REFERENCE: 10338-5US

CURRENT APPLICATION NUMBER: US/03/900,345A

PRIOR FILING DATE: 2001-03-06

PRIOR APPLICATION NUMBER: AU PP-678

PRIOR FILING DATE: 1999-01-06

PRIOR APPLICATION NUMBER: PCT/AU00/00008

PRIOR FILING DATE: 2000-01-07

NUMBER OF SEQ ID NOS: 185

SOFTWARE: Patents Ver. 2.0

SEQ ID NO 125

LENGTH: 238

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: GFP Humanized

OTHER INFORMATION: control

US-09-900-345A-125

Query Match 98.6%; Score 1258; DB 9; Length 238;
Best Local Similarity 98.3%; Pred. No. 6.6e-115;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGELFTGVVPLVVELDGVNGHKFSVSGEGSDATYKLTAFICITGKLPVWPPTL 60

DB 1 MSKGELFTGVVPLVVELDGVNGHKFSVSGEGSDATYKLTAFICITGKLPVWPPTL 60

QY 61 VTTLTYGVOCFSRYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNYKTRAEVKEGGDTLV 120

DB 61 VTTFSGVOCFSRYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNYKTRAEVKEGGDTLV 120

QY 121 NRIELKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180

DB 121 NRIELKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 1 MSKGELFTGVVPLVVELDGVNGHKFSVSGEGSDATYKLTAFICITGKLPVWPPTL 60
DB 1 MSKGELFTGVVPLVVELDGVNGHKFSVSGEGSDATYKLTAFICITGKLPVWPPTL 60
QY 61 VTTLTYGVOCFSRYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNYKTRAEVKEGGDTLV 120
DB 61 VTTFSGVOCFSRYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNYKTRAEVKEGGDTLV 120
QY 121 NRIELKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
DB 121 NRIELKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238
DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

RESULT 12

US-10-221-461-6

Sequence 6, Application US/10221451

Publication No. US20030092902A1

GENERAL INFORMATION:

APPLICANT: Marsh, Donald J.

TITLE OF INVENTION: MELANIN CONCENTRATING HORMONE RECEPTOR

TITLE OF INVENTION: CHIMERIC AND FUSION PROTEINS

FILE REFERENCE: 20652P

CURRENT APPLICATION NUMBER: US/10/221,451

CURRENT FILING DATE: 2002-09-12

PRIOR APPLICATION NUMBER: PCT/US01/08071

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/189,598

PRIOR FILING DATE: 2000-03-15

NUMBER OF SEQ ID NOS: 37

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 6

LENGTH: 238

TYPE: PRT

ORGANISM: Aequorea Victoria

US-10-221-461-6

Query Match 98.6%; Score 1255; DB 9; Length 235;

Best Local Similarity 98.3%; Pred. No. 6.6e-115;

Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGELFTGVVPLVVELDGVNGHKFSVSGEGSDATYKLTAFICITGKLPVWPPTL 60

DB 1 MSKGELFTGVVPLVVELDGVNGHKFSVSGEGSDATYKLTAFICITGKLPVWPPTL 60

QY 61 VTTLTYGVOCFSRYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNYKTRAEVKEGGDTLV 120

DB 61 VTTFSGVOCFSRYPDHKKRHDFPKSAMPEGVVOERTIFFKDDGNYKTRAEVKEGGDTLV 120

QY 121 NRIELKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180

DB 121 NRIELKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180

QY 141 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLEFVTAAGITGHMDELYK 238

RESULT 13

US-09-920-922-4

Sequence 4, Application US/9920922

Patent No. US2002008448A1

GENERAL INFORMATION:

APPLICANT: Miyawaki, Asushu

APPLICANT: Sawano, Asuko

TITLE OF INVENTION: METHOD FOR MUTAGENESIS

FILE REFERENCE: 11283-012001

CURRENT APPLICATION NUMBER: US/09/920,922

CURRENT FILING DATE: 2001-08-02

US-09-920-922-4

Sequence 4, Application US/9920922

Patent No. US2002008448A1

GENERAL INFORMATION:

APPLICANT: Miyawaki, Asushu

APPLICANT: Sawano, Asuko

TITLE OF INVENTION: METHOD FOR MUTAGENESIS

FILE REFERENCE: 11283-012001

CURRENT APPLICATION NUMBER: US/09/920,922

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: JP 2000-237166
PRIOR FILING DATE: 2000-08-04
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 238
TYPE: PRT
ORGANISM: Acetivorea victoria
US-09-920-922-4

Query Match 98.6%; Score 1258; DP 16; Length 238;
Best Local Similarity 98.3%; Pred. No. 6.6e-115;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGSDATYGLTKLTKFCTTCKLPVPWPTL 60
DB 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGSDATYGLTKLTKFCTTCKLPVPWPTL 60
QY 61 VTTLTYGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
DB 61 VTTLTYGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
QY 121 NRLELKGIDFKEDGNILGHKLEYNYNHNYIMADKOKNGIKVNFKTRHNIEDGGVGLAD 180
DB 121 NRLELKGIDFKEDGNILGHKLEYNYNHNYIMADKOKNGIKVNFKTRHNIEDGGVGLAD 180
QY 181 HYQONTPIODGPVLLPDNHYLSTOSALSADPNKRDHNVLLFEFVTAAGTTHGMDIELYK 238
DB 181 HYQONTPIODGPVLLPDNHYLSTOSALSADPNKRDHNVLLFEFVTAAGTTHGMDIELYK 238

RESULT 14

US-09-952-000-1
Sequence 1, Application US/95852409
Patent No. US20020099170A1
GENERAL INFORMATION:
APPLICANT: Osumi, Takashi
APPLICANT: Tsukamoto, Toshio
APPLICANT: Tsukamoto, No. US20020099170A1
APPLICANT: Yamasaki, Masao-Oshi
TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
TITLE OF INVENTION: PROTEINS
FILE REFERENCE: 046124-5005-US
CURRENT APPLICATION NUMBER: US/09/852-000
PRIOR FILING DATE: 2001-03-10
PRIOR APPLICATION NUMBER: JP 024418/1998
PRIOR FILING DATE: 1998-01-23
PRIOR APPLICATION NUMBER: US 09/121,539
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 09/515,655
PRIOR FILING DATE: 2000-07-13
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 1
LENGTH: 238
TYPE: PRT
ORGANISM: Acetivorea victoria
FEATURE:
OTHER INFORMATION: Green fluorescent protein
US-09-852-000-1

Query Match 98.6%; Score 1258; DP 16; Length 238;
Best Local Similarity 98.3%; Pred. No. 6.6e-115;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGSDATYGLTKLTKFCTTCKLPVPWPTL 60
DB 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGSDATYGLTKLTKFCTTCKLPVPWPTL 60
QY 61 VTTLTYGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
DB 61 VTTLTYGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAFVKFEGDTLV 120

QY 121 NRLELKGIDFKEDGNILGHKLEYNYNHNYIMADKOKNGIKVNFKTRHNIEDGGVGLAD 180
DB 121 NRLELKGIDFKEDGNILGHKLEYNYNHNYIMADKOKNGIKVNFKTRHNIEDGGVGLAD 180
QY 181 HYQONTPIODGPVLLPDNHYLSTOSALSADPNKRDHNVLLFEFVTAAGTTHGMDIELYK 238
DB 181 HYQONTPIODGPVLLPDNHYLSTOSALSADPNKRDHNVLLFEFVTAAGTTHGMDIELYK 238
RESULT 15
US-09-899-954B-8
Sequence 8, Application US/9989954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
TITLE OF INVENTION: Target cell or tissue of a plant
FILE REFERENCE: 10338-6US
CURRENT APPLICATION NUMBER: US/09/999-954B
PRIOR FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU PP8077
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: PCT/AU00/0607
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Patent in version 3.1
SEQ ID NO 8
LENGTH: 244
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Len(CTA)6 GFP construct
US-09-899-954B-8

Query Match 98.4%; Score 1256; DP 9; Length 244;
Best Local Similarity 97.9%; Pred. No. 1.1e-114;
Matches 233; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGSDATYGLTKLTKFCTTCKLPVPWPTL 60
DB 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGSDATYGLTKLTKFCTTCKLPVPWPTL 60
QY 61 VTTLTYGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
DB 61 VTTLTYGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
QY 121 NRLELKGIDFKEDGNILGHKLEYNYNHNYIMADKOKNGIKVNFKTRHNIEDGGVGLAD 180
DB 121 NRLELKGIDFKEDGNILGHKLEYNYNHNYIMADKOKNGIKVNFKTRHNIEDGGVGLAD 180
QY 181 HYQONTPIODGPVLLPDNHYLSTOSALSADPNKRDHNVLLFEFVTAAGTTHGMDIELYK 238
DB 181 HYQONTPIODGPVLLPDNHYLSTOSALSADPNKRDHNVLLFEFVTAAGTTHGMDIELYK 244

Search completed: July 16, 2003, 16:30:11
Job time: 30 secs

GenInfo version 5.1.6
Copyright (c) 1993 - 2003 Computer 1.1.

CM protein - protein search, using sw model:

Run on: July 16, 2003, 16:07:32 : Search time 25.6667 seconds
(without alignments)
1:07:396 Million cell updates/sec

Title: us-09-967-301-4

Perfect score: 1236

Sequence: 1 MKSGEETVWVPIVLEDC.....VLSEVFAATTHMDKYS 23P

Scoring table: BKA0062

Gapop 10.0 : Gapext 0.5

Searched: 281224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 281224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 04
Maximum Match 1008
Listing first 45 summaries

Database : PIR73:
1: PIR1:
2: PIR2:
3: PIR3:
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1252	98.1	238	1 J01514	green-fluorescent protein
2	100	7.8	785	2 H72228	hypothetical prote
3	95.5	7.7	887	2 H82590	leucyl-tRNA synth
4	92.5	7.2	534	1 NICDMA	nitrogenase (EC 1.
5	92	7.2	336	2 C64468	hypothetical prote
6	91	7.1	439	2 JH0414	synaptobactin G-655
7	89	7.0	1224	1 E9H0A8	cholesterol complex a
8	88.5	6.9	861	2 H64102	leucine-tRNA ligase
9	86	6.9	531	2 C95338	hypothetical prote
10	87.5	6.9	370	2 E70390	iron-sulfur cofact
11	87	6.8	357	2 G81355	tRNA (tyrosyl-S)-m
12	86.5	6.8	2222	1 A35028	DNA-directed DNA p
13	86.5	6.8	2573	2 D71614	hypothetical prote
14	86	6.7	241	2 AD2052	hypothetical prote
15	85.5	6.7	263	2 S53488	water-stress-induc
16	85.5	6.7	700	1 C1HCH2	galactin (EC 3.4.22
17	85.5	6.7	797	2 T03079	protective surface
18	85.5	6.7	806	2 F64102	protective surface
19	85.5	6.7	860	2 A35082	leucyl-tRNA synth
20	85	6.7	47	2 T27856	hypothetical prote
21	85	6.7	632	2 T05886	DNA-binding protei
22	84.5	6.6	564	2 F81117	ABC-type transport
23	84.5	6.6	655	2 D83917	DNA topoisomerase
24	84.5	6.6	883	2 J05576	inter-alpha-trypsi
25	83.5	6.5	422	2 B24815	calpain (EC 3.4.22
26	82.5	6.5	312	2 C81710	cholesterol reduct
27	82.5	6.5	1134	1 F60234	EGF receptor pr
28	82.5	6.5	1164	1 FC0043	EGF receptor pr
29	82	6.4	470	2 T31349	hypothetical prote

30	81.5	6.4	330	2 C90208	hypothetical prote
31	81.5	6.4	513	2 A99552	oligodehydrogenase
32	81.5	6.4	536	2 C70031	hypothetical prote
33	81.5	6.4	788	1 GDLVHH	DNA-directed DNA p
34	81.5	6.4	2518	2 A12140	polyketide synthas
35	81	6.3	422	1 BMS72Y	synaptobactin II -
36	81	6.3	547	2 H89988	hypothetical prote
37	81	6.3	574	2 J04930	S-layer protein pr
38	80.5	6.3	329	2 G84245	NADH dehydrogenase
39	80.5	6.3	393	2 C64613	conserved hypother
40	80.5	6.3	752	1 KRP0C1	proprotein convert
41	80.5	6.3	1269	2 AE1055	probable exported
42	80	6.3	353	2 F84941	dimethylglycerol d
43	80	6.3	578	1 I40794	probable oxysterol
44	80	6.3	760	2 E84715	hypothetical prote
45	80	6.3	764	2 T25012	hypothetical prote

ALIGNMENTS

RESULT 1

J01514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence,revision 03-Dec-1999 text_change 23-Mar-2001
C:Accession: J01514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prasher, D.C.; Gorman, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green fluorescent protein.
A:Reference number: J01514; MUID:92175527; PMID:1347277
A:Accession: J01514
A:Molecule type: DNA
A:Residues: 1-307, S, 109-238 <PRA1>
A:Cross-references: GB:M62654; NID:q155662; PIDN:AAA27722.1; PID:q155663
A:Accession: J01514
A:Molecule type: mRNA
A:Residues: 1-59, F, 101-140, L, 142-218, V, 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:q155660; PIDN:AAA27721.1; PID:q155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64, 74-122, 132-151, 154-183, 185-200 <PRA3>
R:Frederick, S.; Tsai, F.L.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, Q, 26-156, P, 158-171, K, 173-238 <INO>
A:Cross-references: GB:L29345; NID:q606303; PIDN:AAA58246.1; PID:q606384
R:Watkins, J.N.; Campbell, A.K.
Submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13, V, 15-24, Q, 26-44, N, 46-153, G, 155-156, P, 158-171, K, 173-227, F
A:Cross-references: EMBL:X83959; NID:q634008; PIDN:CAA58789.1; PID:q634009
A:Experimental source: clone gfp
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24, Q, 26-29, R, 31-84, L, 85-153, G, 155-156, P, 158-171, K, 173-228, F
A:Cross-references: EMBL:X83960; NID:q634010; PIDN:CAA58790.1; PID:q634011
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65892; PDB:1GFI.
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-73, F, 78
A:Note: engineered sequence based on J01514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58953; MUID:98294543; PMID:9631087

A:Contents: annotation: X-ray crystallography, 1.3 angstroms
C:Comment: This protein is excited by the photoprotein reaction (see P16049FW) emitting
C:Comment: The chromophore of this protein is formed by modification of Ser-64hydro-Tyr-
C:Genetics:
A:Gene: GFP
A:Introns: 69/3; 167/3
A:Superfamily: green-fluorescent protein
C:Keywords: chromoprotein; luminescence
F:65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 98.1%; Score 1252; DB 1; Length 238;
Best Local Similarity 97.1%; Pred. No. 2.4e-93;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFGVVPILVELDSDVNGHKFSVSGEGGDAATYKGLLAKFTCTTKLVPPVPTL 60
DB 1 MSKGEELFGVVPILVELDSDVNGHKFSVSGEGGDAATYKGLLAKFTCTTKLVPPVPTL 60
QY 51 VITLTVGVOCEGRYPDHMKRHDPEKSAMPEGVVQERTIFKDGNGYKTPAEVKFGDILV 120
DB 51 VITLTVGVOCEGRYPDHMKRHDPEKSAMPEGVVQERTIFKDGNGYKTPAEVKFGDILV 120
QY 121 NRIELKGDIFKEDONILGHKKLEYNVNSHNYVINALKQNTYKVNKKEPNTEDSGVOLAD 180
DB 121 NRIELKGDIFKEDONILGHKKLEYNVNSHNYVINALKQNTYKVNKKEPNTEDSGVOLAD 180
QY 181 HYQONTPIGDGVLLPDNIYLTQSALSQDKPKRKHVLLFVTAAGTTHOMFELYK 236
DB 181 HYQONTPIGDGVLLPDNIYLTQSALSQDKPKRKHVLLFVTAAGTTHOMFELYK 236

RESULT 2
R72228
hypothetical protein TM1624 - Thermotoga maritima (strain MS94)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2003
R:Nelson, K.E.; Clayton, R.A.; Gill, S.P.; Oxiom, M.M.; Dodson, R.J.; Haft, D.B.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, K.S.; Phillips, G.A.; Richardson, P.
C.M.

Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genomic seq
A:Reference number: A72228; MID:99287416; PID:1060571
A:Accession: H72228
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-785 <APN>
A:Cross-references: GB:AE001502; GB:AE000512; NID:q10494; PIN:AM 4490.1; L0010494-219
A:Experimental source: strain MS94
C:Genetics:
A:Gene: TM1624

Query Match 7.8%; Score 100; DB 2; Length 785;
Best Local Similarity 20.9%; Pred. No. 1.0;
Matches 40; Conservative 27; Mismatches 54; Indels 6; Gaps 0;

QY 2 SKGRLEFGVVPILVELDSDVNGHKFSVSGEGGDAATYKGLLAKFTCTTKLVPPVPTL 41
DB 15 NEGRSPGEGVPGVQAD-----LVKGLDHPHVGK- 46
QY 52 TLTLYGVOCEGRYPDHMKRHDPEKSAMPEGVVQERTIFKDGNGYKTPAEVKFGDILV 121
DB 52 TLTLYGVOCEGRYPDHMKRHDPEKSAMPEGVVQERTIFKDGNGYKTPAEVKFGDILV 121
QY 122 RIELKGDIFKEDONILGHKKLEYNVNSHNYVINALKQNTYKVNKKEPNTEDSGVOLAD 181
DB 122 RIELKGDIFKEDONILGHKKLEYNVNSHNYVINALKQNTYKVNKKEPNTEDSGVOLAD 181
QY 182 YQONTPIGDGP 192
DB 182 YQONTPIGDGP 192
QY 192 YQONTPIGDGP 192
DB 192 YQONTPIGDGP 192

RESULT 3
E82590
leucyl-tRNA synthetase XF2176 (imported) - Xylella fastidiosa (strain 945c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: E82590
A:Anonymous, The Xylella fastidiosa Consortium of the organization for Nucleotide Seq
Nature 405, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59128 below
A:Accession: E82590
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-887 <Sim>
A:Cross-references: GB:AE004031; GB:AE003849; NID:q9107309; PID:AA84975.1; GSPB:GN
A:Experimental source: strain 945c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Accacio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Cartao, L.E.A.; Cartao, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.C.S.
submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Fr
C.D.; Junqueira, M.L.; Komper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.F.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., K.A.; da Silv
M.; Iwakura, M.H.; Vahlada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.C.
A:Reference number: A59128
A:Contents: annotation
C:Genetics:
A:Gene: XF2176
C:Superfamily: leucine-tRNA ligase

Query Match 7.7%; Score 98.5; DB 2; Length 887;
Best Local Similarity 23.2%; Pred. No. 2.8;
Matches 46; Conservative 27; Mismatches 68; Indels 57; Gaps 10;

QY 49 ITGKLPVPMPLVILTYGVQCFRSYDPDHMKRHDPEKSAMPEGVVQERTIFKDGNGY-- 106
DB 329 INQLPV-WYANVIVMAYVIGAVMAYVPGHQRQFF--ANKYGLPIQVIA:KEPKNLE 385
QY 107 -----KTRAEVKFGDILVNRILKIDGFKEDONILGHKKLEYNVNSHNYV 152
DB 111 -----KTRAEVKFGDILVNRILKIDGFKEDONILGHKKLEYNVNSHNYV 152
QY 182 STWEDPVWEDWYADKTR--EEF---LINSAEEDNIEYQDAFEVLAERFE--- 429
DB 153 MADKQKNG:IKVNEKIRNNTEDSGVOLADCHYQNTPI-----GNGVPLFEN 196
DB 430 ---RQGGGGRVWYHLP- --DWGNSRQRYWGCPIPVITCTCGGAVPVFELQIPVILFEN 482

QY 199 -HILSTVSALSQDKDFEKA 215
DB 483 VAFSGTSPKIDPFWEK 500

RESULT 4
NICLMA
nitrogenase (EC 1.18.6.1) molybdenum-iron: protein alpha chain [validated] - Clostridi
N:Alternate names: dinitrogenase alpha chain; nitrogenase component I alpha chain
C:Species: Clostridium pasteurianum
C:Date: 01-Sep-1981 #sequence_revision 31-Dec-1991 #text_change 19-Jan-2001
C:Accession: S07389; I40816; A00545
R:Wang, S.Z.; Chen, J.S.; Johnson, J.L.
Nucleic Acids Res. 15, 3935, 1987

A:Title: Nucleotide and deduced amino acid sequences of nitro encoding the alpha-subun
A:Reference number: S07389; MID:97231095; PMID:3473447
A:Accession: S07389
A:Molecule type: DNA
A:Residues: 1-534 <WAN>
A:Cross-references: EMBL:Y00155; NID:q40583; PID:CAA68349.1; PID:q580995
R:Chen, K.C.


```

Db 97 GLCTCKKCCCKKKKXKKEKXK-----KXHNK-----DVKSGDNQED 136
QY 117 --DTLVNRLEKIGID--KEUGNT--LGHKLEYNYSN----- 146
Db 139 DAEATGTLTGEDKAEKAEKCKKTOPSDYDFGKQNLITLVQMAELPALDPAWTSTSEY 196
QY 149 -NVYIMADQKN-GIKVN-----FKRHNIEDSGVOLA-----DHYQGNTP 187
Db 199 VKVELLPKKKKRYKVKCKTKTLNPTFNSEFVKVYQ-ELSGKTKIMAVYD-DPFSKIDC 257
QY 188 IGGCVLLPD-----NHVLTQSALSKDKNKEDHMLVLEFYTAG 228
Db 258 IGGVTLMTKYDLOGGLEWRDLSEAKBEPEKJGDCICTSLRYPTAG 305

RESULT 7
EXHUAH
C:contamer complex alpha chain homology - human
N:Alternate names: HEP-COP; xenopsin homo-co; xenopsin-related peptide precursor
C:Species: xenin; 25; xenopsin-related peptide
C:Date: 10-May-1996 #sequence_revision 08-Nov-1996 #text_change 01-Jul-2000
C:Accession: J04668; A44317
R:Chow, V.T.X.; Osek, H.H.
Gene 169, 223-227, 1996
A:Title: HEP-COP, a novel human gene whose product is highly homologous to the alpha-subunit of the xenin receptor
A:Reference number: J04668; MUID:96194866; PMID:8647451
A:Accession: J04668
A:Molecule type: mRNA
A:Residues: 1-1224 <COP>
A:Cross-references: GB:U24105; NID:q1638823; PIDN:A44317; PID:q1638823
A:Experimental source: Hep3B hepatocellular carcinoma cell
R:Farrie, G.B.; Hanscher, G.; Kuslek, A.; Meyer, H.P.; Metzger, T.W.
J. Biol. Chem. 267, 22305-22309, 1992
A:Title: Identification of xenin, a xenopsin-related peptide, in the human gastrin receptor
A:Reference number: A44317; MUID:9305415; PMID:1424561
A:Accession: A44317
A:Molecule type: protein
A:Residues: 1-25 <PEJ>
A:Experimental source: gastric mucosa
A:Note: sequence extracted from MORT backbone (NCBI:U17024)
C:Genetics:
A:Gene: GDB:COPA; HEP-COP
A:Cross-references: GDB:4642737; OMIM:601924
A:Map position: 1q23-1q25
C:Superfamily: yeast coatomer complex alpha chain; WD repeat homology
C:Keywords: duplication; hormone; plasma; stomach
F:1-25/Product: xenin 25 #status experimental <XNP>
F:15-38/Domain: WD repeat homology <WDL>
F:17-25/Product: xenopsin-related peptide #status predicted <XRP>
F:47-80/Domain: WD repeat homology <WR2>
F:89-122/Domain: WD repeat homology <WR3>
F:131-164/Domain: WD repeat homology <WR4>
F:201-234/Domain: WD repeat homology <WR5>
F:245-278/Domain: WD repeat homology <WR6>

Query Match 7.0%; Score 89; DB 1; Length 1224;
Best local similarity 25.9%; Pred. No. 27;
Matches 41; Conservative 26; Mismatches 59; Indels 32; Gaps 7;

QY 90 EGYVQERTIFFKDCGNYKTRAEVKEFGDILVNRIE-LKSIDFKEDGNILGHKLEYNYSN 146
Db 692 KGYPEVALHFVDE---KTRFLA-ECGNIE-ALAAKALDKCKNCKWELGSHVALLQGNHQ 688
QY 149 NVYIMADQKN-GIKVN-----FKRHNIEDSGVOLA-----DHYQGNTP 197
Db 689 NVYIMADQKN-GIKVN-----FKRHNIEDSGVOLA-----DHYQGNTP 197
QY 198 NHVLTQSALSKDKNKEDHMLVLEFYTAGIHWDE 235
Db 742 ---VSEVRVLLKNCCKCKS-----LAYLTAA---THOLDE 769

```

```

RESULT 8
H64102
C:leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain AG KW2C)
N:Alternate names: leucyl-tRNA synthetase
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
C:Accession: H64102
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage,
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, S.M.; Weidman,
; D.M.; Brandon, R.C.; Fire, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: H64102
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-861 <TIGR>
A:Cross-references: GB:U32774; GR:A2023; NID:q1573942; PIDN:AG22581.1; PID:q1573943
C:Genetics:
A:Gene: leus
C:Superfamily: leucine-tRNA ligase
C:Keywords: aminocyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 6.9%; Score 88.5; DB 2; Length 861;
Best local similarity 24.3%; Pred. No. 19;
Matches 45; Conservative 26; Mismatches 67; Indels 47; Gaps 9;

QY 49 TTKLPPVPTLVLEIYVVOGCFSRVFDHMKRDETFKSAMPVGVVETIFFKLGNYK 128
Db 314 TPKLPI-WYANFVLMYHSGAYMAVPAHQDQD--FFAQKYSLPKQVIAPLADRIDL 370
QY 109 KAFYKRGELTVNRLEKIGIDFKEDGNILGHKLEYNYSNHYIMADQKN-GIKVN 164
Db 372 TKAFVEHSGKLVNSDEFGKNF--DGNFNG-----IAKLEKLVGGRQVN 414
QY 165 FKTRH-----NTEDGSQLADIYQNTPTGGGPFVLGFENHY-SPQSAL 227
Db 415 YRLDMQVSRQKYGADIPKLTLSNVDVPA-----PMEDLPILPDVGVGVKSP: 467
QY 205 SKQPN 212
Db 468 NACPN 472

RESULT 9
Q15338
C:hypothetical protein Sma103 [imported] - Sinorhizobium meliloti (strain 102) masap1
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: Q15338
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, G.; Abola, A.P.; Barley-Hubler, P.; B.
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.G.; Surzycki, R.; Wells, D.H.; Yeh, K.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: Q15338
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-531 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK65269.1; PID:q14523721; GSPDB:GNC0165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barley-Hubler,
; Pella, D.; Chait, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, G.; Lelau
; hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:

```


A:Gene: SMal131
A:Genome: plasmid

Query Match 6.3% Score 88: DB 2: Length 532
Best Local Similarity 20.4% Pred. No. 11:
Matches 61: Conservative 29: Mismatches 27: Indels 122: Gaps 14:
QY 6 ELFTGVVPLVLEGGVNV-CHKFSVSGEGBDATTYKGLDKFKIGIKGLVNPVLTATL 64
DB 176 EUTGSPRLFLFSGLISPRHL-----GGP- YAKAFSGWGVDS 216
QY 65 TYGVQCFSPYPMKKHH-----DFKSAFEG-----YVLFK-----IFPK----- 162
DB 217 TYGNTEREADAAARPHILRSEVLTAHPNGALLSPSPAFVFTQSLIDVHVMESAMP 276
QY 203 -----DGNVKTAEVKF-----EGDTLVNKEIKGLKKE----- 142
DB 277 KCFILDSPLATRASIFRHHARELNGALVATSKSNVPTFAPQSHVLLIFSH 336
QY 233 -----DGNLGHKLE-FYNSHNVYIMADKKNP-----TAVN 154
DB 337 VIAASGCEAGRIERIKWLRDEATVLLVGYDANCLLELLESATLVVAGQDIPYR 466
QY 165 FKTH-NIEG---GVLADH---GNIFIGSPV-----GFGNHYL 201
DB 397 AR-PEKLVYSHADGSLADKWRVARGPISAGVNVHGHENIRIFERLSTFQWED 456
RESULT 1C
E70390
Iron-sulfur cofactor synthesis protein nifs Aquifex aeolicus
N:Contains: L-cysteine sulfurtransferase (EC 2.8.1.1)
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 *sequence_revision 08-May-1998 *text_change 07-Dec-1999
C:Accession: E70390
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Leroch, A.L.; Graham, L.E.; Oj
V.

Nature 392, 353-358, 1995
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; WMD:98156666; PMID:9537329
A:Accession: E70390
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-370 <AO>
A:Cross-references: GB:AE000720; NID:q2983529; P1DN:AA037111; P1D:q2984536; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:

A:Gene: nifs1
C:Superfamily: nitrogen fixation protein nifs
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase
F:195/Binding site: pyridoxal phosphate (lys) (cofactor) *status predicted
F:319/Active site: cys (cysteine persulfide intermediate) *status predicted

Query Match 6.3% Score 67.5: DB 2: Length 370:
Best Local Similarity 25.4% Pred. No. 7:
Matches 49: Conservative 30: Mismatches 9: Indels 23: Gaps 6:
QY 3 KGEELFTGVV-----PHIVELD---GVNKHKF-SVSEH---EGVATYGLKLEKFIN 14
DB 164 KGVPLLDVAQAGCKPIELKNISYATFESHKPKHAKGSPFATSGEATPEPLIVGGGG 224
QY 50 TGKLP-----VPMPLVTHITTYGVQCFSPYPMKKR-HDFKSAFPPYVQGNIFPKD 109
DB 224 NGKRSUTENVVGLSLAKALEIIVSNFSRYOGLKKLDFLFENLIEA-LPDAQIVKQKA 284
QY 104 GNVKTRAV---AFEGDTLVNKLKLDIFKFGNGLGHKLENYNSHNVYIMADKKN 160
DB 283 ERSPSISGVMPAPFGAIVNKLSEKGYCSGSAGLSDFYFENKMLNMFSEKALAK 342
QY 161 IKVNFKRRHIND 173
DB 343 VRPSFGLLNKEE 355

RESULT 11

GB1355
RNA (cyticil-5)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter j
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 *sequence_revision 31-Mar-2000 *text_change 03-Jul-2002
C:Accession: GB1355
C:Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chit
C.W.; Ouail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Bai
Nature 403, 665-669, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: AB1250; NCID:20150912; PMID:10688204
A:Accession: GB1355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <PA>
A:Cross-references: GB:AL11159; NID:q6966128; P1DN:CAH73596.1; P1D:q6966
A:Experimental source: serotype O2, strain NCIC 11158
C:Genetics:
A:Gene: trmA; Cj0831c
C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 6.8% Score 87: DB 2: Length 357:
Best Local Similarity 24.8% Pred. No. 8:
Matches 30: Conservative 16: Mismatches 49: Indels 54: Gaps 5:
QY 79 KRHDFKSAFEGYVQENTIFSKDIGNYKTRAEVKF---EGDTLV-----NPIELKG 127
DB 14 EKHSFKKFKFEFFYTKPKFKASKDKHYKRAELSFYHENDILFYAMHDKSKKYLLEY 73
QY 128 DFKED-----GNLGHKLENYNSHNVYIMADKKNKIKVFKIRHNE 172
DB 74 LDFADEKICAEERILEYLRQDNK-LKEKI-----EGVEFTIKGE--LSITLIVKRIE 125
QY 173 D 173
DB 126 D 126

RESULT 12

A36028
DNA-directed DNA polymerase (EC 2.7.7.7) II catalytic chain - yeast (Saccharomyces ce
N:Alternate names: DNA-directed DNA polymerase II chain A; protein N0825; protein YNL
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 *sequence_revision 10-Sep-1999 *text_change 21-Jul-2000
C:Accession: A36028; B36028; S60919; S63235; S65121
R:Morison, A.; Araki, H.; Clark, A.B.; Hamdake, R.K.; Sugino, A.
Cell 62, 1143-1151, 1990
A:Title: A third essential DNA polymerase in Saccharomyces cerevisiae.
A:Reference number: A36028; NCIT:90381771; PMID:2169349
A:Accession: A36028
A:Molecule type: DNA
A:Residues: 1-2222 <MO>
A:Cross-references: GB:M60416; GB:M36724; NID:q171408; P1DN:AAH8711.1; P1D:q171409
A:Accession: B36028
A:Molecule type: protein
A:Residues: 1214-1216 'X', 1218-1221 <MO>
R:Sen-Gupta, M.; Lyck, R.; Fiebig, U.; Niedenthal, R.N.; Hegemann, J.H.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV
A:Reference number: S60909
A:Accession: S60919
A:Molecule type: DNA
A:Residues: 1-2221 <SEN>
A:Cross-references: EMBL:X52494; NID:q1045236; P1DN:CAH63235.1; P1D:q1045247
R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fiebig, U.N.; Hegemann, J.H.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63235
A:Accession: S63235
A:Molecule type: DNA
A:Residues: 1-2222 <SEN>
A:Cross-references: EMBL:271538; NID:q1302316; P1DN:CAA6169.1; P1D:q1302317; GSPDH:G
A:Experimental source: strain S288C

R/Son-Gupta, M.; Dyck, R.; Field, G.; Niedenthal, R.K.; Regenstein, J.R.
Yeast 22: 505-514, 1996
A:Title: The sequence of a 24 kb segment from the left arm of chromosome XIV from *Saccharomyces cerevisiae* strain 1024-11A
A:Reference number: S65111; MID:96310651; PMID:89740425
A:Accession: S65121
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2221 <SEF>
A:Cross-references: EMBL:X82434; NID:91045236; PIR:CAA63435.1; PMID:8045247
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C:Genetics:
A:Gene: SGD:POL2; DUN2: MFS:YNL262w
A:Cross-references: SGD:S0005206; MIPS:YNL262w
A:Map position: 14L
A:Superfamily: DNA-directed DNA polymerase II
C:Keywords: DNA binding; nucleotidyltransferase; nucleus; zinc finger

Query Match 6.8% Score 86.5; DB 2; Length 2222;
Best Local Similarity 28.2%; Pred. No. 97;
Matches 37; Conservative 14; Mismatches 45; Indels 31; Gaps 7;

QY 53 LPVWP-ILVTTILYGVQCSRPDHN-----KREIFFKSAMPGV-----SEPTI 96
DB 883 LKRSFTEIFELANGKKLYSPCSMLNYPVSEKFTNEGYELKDFLNLYTHSENTI 942
QY 99 FFKDDGNKTR--AEVFKFGDTLVNR-----IEKAGTDKFGNGLGHKLEYN 146
DB 943 EFEVGGYKAMILPSSKEEKGGIKKRYAVNHLGSLAEKSGEKRRAGE--QILKNVQ 989
QY 147 S--HNVTIMAD 155
DB 1000 SDIFKVPLEGD 1010

RESULT 13
D71614
Hypothetical protein PF0450c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: D71614
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Arayind, L.; Koonin, E.V.;
Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, G.; Smith, R.;
Science 282: 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum
A:Reference number: A71600; MID:95221743; PMID:960461
A:Accession: D71614
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2573 <QAR>
A:Cross-references: GB:AE001396; GS:AE001362; NID:93451268; PIR:AA71382.1; PMID:4384719
A:Experimental source: clone 307
C:Genetics:
A:Gene: PF0450c

Query Match 6.8% Score 86.5; DB 2; Length 2573;
Best Local Similarity 26.2%; Pred. No. 112; Gaps 257;
Matches 34; Conservative 29; Mismatches 54; Indels 14; Gaps 5;

QY 93 VQERTFFKQ--CGNPKTRAEVFEGLTVNRIELKGLDKEGNTLGHKLEYN--YKSE 148
DB 126 LKRSITLLCKD:KSGSNPMDTSLFKDMDYQKECK--DEKSKSKKKNKRVNFTYNNK 26-

QY 149 NVYINAKQKNG:KVNFKIRENTEDGSGVLADHYQONTPTGQVLPONHLYSTSAIS 20F
DB 184 NLRIKENKKKQKREKKN:HNNDNNK----IYKNI----DKTHILDNNVVELINCIN 246

QY 209 KDPNKKRDM 218
DB 237 TYLKRBDYM 246

RESULT 14
AD2052

Hypothetical protein cl111970 (imported, - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2052
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, I.; Sasamoto, S.; Watanabe, A.; Iriki
Nakazaki, N.; Shirao, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata
DNA Res. 8: 205-213, 2001
A:Title: Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium
A:Reference number: AB1807; MID:21595285; PMID:11759840
A:Accession: AD2052
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-261 <KOR>
A:Cross-references: GB:BA000619; PIR:BA073669.1; PIR:SI7131040; SPS:05-0306-179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: cl11970

Query Match 6.7% Score 86; DB 2; Length 281;
Best Local Similarity 24.4%; Pred. No. 72;
Matches 33; Conservative 26; Mismatches 54; Indels 22; Gaps 6;

QY 106 YKTRAEVFEGLTVNRIELKGI----DFKEDGNTLGHKLEYNHNVINACKQKNGI 161
DB 46 FKARS-LOSSDPIILNPKYNNLONRPPKKDDEYIVHAEGYKYE-----SSKRKHSI 99
QY 162 KYNFKI--HNTEGSGVLADH-----YQNTPIGQG--VLLPNHNYLS--GSALSK 209
DB 200 KPGSLQERGIIVNQVLQDLNVLTPDEPIOTDPTPSQONTPLANRCKPISUPKTIK 159
QY 220 DPKKRDHNVILEFV 224
DB 160 QCKELSKDKIGERY 174

RESULT 15
S53486
water-stress-inducible protein DS2 - Chaco potato
C:Species: Solanum chacoense (Chaco potato)
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C:Accession: S53486
R:Silhavy, D.; Melvaquer, G.; Barta, E.; Bantialva, Z.;
Plant Mol. Biol. 27: 587-595, 1995
A:Title: Isolation and characterization of a water-stress-inducible cDNA clone from *S.*
A:Reference number: S53488; MID:95201251; PMID:7994021
A:Accession: S53486
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-263 <SIL>
A:Cross-references: GB:U12439; NID:9607904; PIR:AA06052.1; PIR:9607905

Query Match 6.7% Score 85.5; DB 2; Length 263;
Best Local Similarity 22.3%; Pred. No. 73;
Matches 46; Conservative 21; Mismatches 86; Indels 51; Gaps 6;

QY 3: GEGEDATYVGLTLKFICTIGKLPVWPTLVITLYGVQCSRPFLMKRHLFEKSAPE 90
DB 4: GSEKTSYSGKTSYSGDPTYE-----KTTFGDD--NKYGEYSYGLTYGKPT 79

QY 91 GY-----VQERTFFKQSNKTRAEVFEGLTVNRIELKGLDKEGNTLGHKLEY-- 143
DB 80 SYGGDNTYGEKTSYSGKDNKYGEKTSYSG-----DGNKYGEKTSYSG 123

QY 144 -NYSNHYINADKQKNGIKVNFKIRHNIEDGGVLADHYQONTPTGQVLLPNHLYLS 202
DB 124 SYGKPSYSGGSDNKYGEKTSYSGNEESYGGVGGETTNYEN-----E 167

QY 203 TGSALSKPNEKRGHVLLEFVTAAG 228
DB 169 SETKTSYKBEKHKHLEEGGLG 193

Search completed: July 16, 2003, 16:14:40
Job time : 21.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM protein - protein search using sw model

Run on: July 16, 2003, 16:11:23 : Search time 10.6667 seconds

(with 3.0 alignment)
926,441 hits (for all) of 1409/500

Title: US-09-967-301-4

Perfect score: 1276

Sequence: 1 MSKGEELPTGVVPLVLELQK.....VLLERVTAAGLTHQHEUVE 236

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112692 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 12582

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1 SwissProt_40; *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1258	98.6	238	1	GFP_ARQV2
2	98.5	7.7	879	1	SYL_XYLEA
3	92.5	7.2	533	1	NIFD_CLOPA
4	92	7.2	336	1	YD48_METJA
5	91	7.1	439	1	SV62_D_SOM
6	89	7.0	1224	1	COPA_HUMAN
7	88.5	6.9	861	1	SYL_HAFIN
8	87	6.8	357	1	TRMA_CAMEL
9	86.5	6.8	2222	1	DPUE_YEAST
10	86	6.7	1224	1	COPA_BOVIN
11	85.5	6.7	700	1	CAN2_HUMAN
12	85.5	6.7	795	1	D152_HAFIN
13	85.5	6.7	797	1	D151_HAFIN
14	84.5	6.6	324	1	CAN2_PIG
15	84.5	6.6	793	1	D153_HAFIN
16	84.5	6.6	886	1	ITH4_MESAR
17	83.5	6.5	422	1	CAN2_RABIT
18	83.5	6.5	533	1	CP61_CANGA
19	82.5	6.5	312	1	TRXS_CHLMO
20	82.5	6.5	1164	1	RNG_STRAG
21	81.5	6.4	890	1	PRYL_SUJMO
22	81.5	6.4	613	1	PRPE_MUSFO
23	81.5	6.4	786	1	CPOL_HPHE
24	81	6.3	422	1	SVT2_RAT
25	81	6.3	874	1	SLAP_RAC11
26	80.5	6.3	393	1	MT04_HAPPY
27	80.5	6.3	538	1	GABE_RAT
28	80.5	6.3	722	1	NECL_RAT
29	80	6.3	354	1	RIS7_RVCA1
30	80	6.3	774	1	AMV2_SCHO
31	80	6.3	953	1	YNM7_YEAST
32	79.5	6.2	312	1	TRXS_CHLTR
33	79.5	6.2	700	1	CAN2_MOUSE

34	79	6.2	504	1	YC03_KLEPN	048469 Klebsiella
35	78.5	6.2	860	1	SYL_ECOLI	007813 Escherichia
36	78.5	6.2	887	1	ETH3_RAI	Q63416 Rattus norv
37	78.5	6.2	941	1	GUN_BACS6	P19424 Bacillus sp
38	78	6.1	421	1	SY11_MOUSE	P46096 Mus musculu
39	77.5	6.1	217	1	EXPT_ERWCA	P33882 Erwinia car
40	77.5	6.1	269	1	PENK_MOUSE	P22095 Mus musculu
41	77.5	6.1	385	1	VGLW_YEAST	P51210 Saccharomyc
42	77.5	6.1	461	1	PSBC_CYAPA	P48114 Cyanophora
43	77.5	6.1	558	1	ADAS_HUMAN	000115 Homo sapien
44	77	6.0	373	1	BPAC_PSEAE	P21175 Pseudomonas
45	77	6.0	490	1	GLYC_CANAL	Q13426 Candida alb

ALIGNMENTS

Result 1	US-09-967-301-4	STANDARD	REF: 238 AA
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		
45	GLYC_CANAL		
34	GFP_ARQV2		
35	SYL_ECOLI		
36	ETH3_RAI		
37	GUN_BACS6		
38	SY11_MOUSE		
39	EXPT_ERWCA		
40	PENK_MOUSE		
41	VGLW_YEAST		
42	PSBC_CYAPA		
43	ADAS_HUMAN		
44	BPAC_PSEAE		

[7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE:90234303; Pubmed:10220315;
RA Elsiger M.A., Wachter R.M., Hanson J.T., Sullivan K.L. Localization of
PT "Structural and spectral response of green fluorescent protein
variants to changes in pH."
RL Biochemistry 36:5296-5301(1999).
CC -!- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSFER THE
CC BLUE CHEMILUMINESCENCE OF THE PROTEIN FLUORESCIN INTO GREEN
CC FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCIN IN VIVO UPON
CC RECEIVING ENERGY FROM THE CA(2+) ACTIVATED PROTEIN AS20-BIR,
CC ASCRAPS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
CC ASCRAPS PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
CC PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.
CC -!- SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: PROTOCYTES.
CC -!- PTM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED
CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.
CC -!- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
CC chimeric proteins of GFP linked to other proteins where it
CC functions as a fluorescent protein tag. GFP tolerates N-and C-
CC terminal fusion to a broad variety of proteins. It has been
CC expressed in bacteria, yeast, slime mold, plants, drosophila,
CC zebrafish, and in mammalian cells. As a noninvasive fluorescent
CC marker in living cells, it allows for a wide range of applications
CC where it may function as a cell lineage tracer, reporter of gene
CC expression, or as a measure of protein-protein interactions.
CC -!- DATABASE: NAME-protein Spotlights.
CC NOTE: Issue 11 of June 2001.
WWW="http://www.expasy.org/spotlights/articles/spl1011.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is not so
modified and this statement is not removed, adapted by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/anno/ncor/>
or send an email to license@sib-sib.ch).

EMBL: M62654; AAA27722.1; .
EMBL: M62653; AAA27721.1; .
EMBL: L29345; AAA58246.1; .
PDB: 1GFL; 11-JAN-97.
PDB: 1EMA; 08-NOV-96.
PDB: 1EMK; 16-JUN-97.
PDB: 1EMC; 20-AUG-97.
PDB: 2EMC; 20-AUG-97.
PDB: 1EMF; 20-AUG-97.
PDB: 1EMF; 20-AUG-97.
PDB: 1EMG; 12-MAY-99.
PDB: 1EMK; 20-AUG-97.
PDB: 1EML; 20-AUG-97.
PDB: 1EMM; 20-AUG-97.
PDB: 2EMM; 20-AUG-97.
PDB: 2EMO; 20-AUG-97.
PDB: 1BFP; 07-JUL-97.
PDB: 1BFP; 28-OCT-98.
InterPro: IPR007866; Green_fl_protein.
Pfam: PF01353; GFP; 1.
PRINTS: P301229; GELFORESCENT.
ProDom: P0013756; Green_fl_protein; 1.
KW Luminescence; 3D-structure.
FT SITE 65 67 MODIFIED TO FORM THE CHROMOPHORE.
FT VARIANT 100 100 F -> Y.
FT VARIANT 108 108 T -> S.
FT VARIANT 141 141 T -> S.
FT VARIANT 219 219 V -> I.
FT VARIANT 25 25 H -> Q (IN REF. 2).
FT CONFLICT 157 157 Q -> P (IN REF. 2).
FT CONFLICT 172 172 P -> K (IN REF. 2).
SEQUENCE 238 AA; 26386 MW; EA5A6F21E9F6E05 CRC64.

Query Match: 95.6%; Score 1258; DS 1; Length 248.
Best Local Similarity: 95.3%; Pred. No. 6,9e+00;
Motifs: 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY : MSKGEELFGVWPIVLELGGVNGVNGHFKFSVSGRGGATYKGLZIKFKICTGKLPYEWPI 60
DL : 1 MSKGEELFGVWPIVLELGGVNGVNGHFKFSVSGRGGATYKGLZIKFKICTGKLPYEWPI 60
QY : 61 VTLLTGWQCFESYPPDHMKRDEFFKSNPGEYVQVRIFFKDSNKKYKAEYKFESEITV 120
DB : 61 VTTFSGVQCFESYPPDHMKRDEFFKSNPGEYVQVRIFFKDSNKKYKAEYKFESEITV 120
QY : 121 KRLEKSGIDFKKGNIIHGHKLEYNKSHNVYVWADQKNGIKVNFKIRNIESGVQLAD 180
DB : 121 KRLEKSGIDFKKGNIIHGHKLEYNKSHNVYVWADQKNGIKVNFKIRNIESGVQLAD 180
QY : 181 HYQNTPIGCPVLLPNHYLSTQSALSADSKOPNKRQHMVLEVTANGITHGMELEYK 238
DB : 181 HYQNTPIGCPVLLPNHYLSTQSALSADSKOPNKRQHMVLEVTANGITHGMELEYK 238

RESULT 2
SYL_XYLPA STANDARD: PKT: 879 AA.
AC QPBG8:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine-tRNA ligase) (LeuRS).
GN LEUS OR XF2176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN 1;
RE SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE:20355717; PubMed:10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Avaronza R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carlet H.,
RA Colauto N.H., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dary H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraça J.S., França S.G., Franco M.C., Frohme M., Furlan L.R.,
RA Gahner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Heheisel J.D., Junqueira M.L., Kemper F.L., Kishima J.,
RA Krieder J.E., Kuramae E., Laigret F., Lambais M.R., Leite R.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado S.A.,
RA Machado M.A., Madeira A.M.S.N., Madeira H.M.P., Marino C.L.,
RA Marques M.V., Martins E.A., Martins E.M.P., Matsukuma A.Y.,
RA Merck C.F.M., Miranda E.C., Miyaki C.V., Monteiro-Vitorello C.E.,
RA Moon L.H., Naqai M.A., Nascimento A.L.T., Netto L.F.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes J.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
RA Peixoto R.R., Pereira G.A.D., Pereira H.A.Jr., Pesquero C.B.,
RA Quadric R.R., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.C., de Souza A.A.,
RA de Souza A.P., Terenzi M.P., Truffi D., Tsai S.M., Tsukano M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa";
RA Nature 406:151-159(2000).
PL -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) -> AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -


```

--53 MACK-----QANG-KVNEK-RENT-EDGCVGLADHYCCTDPIJDSPLYLSP-DNEY 206
      ||| | :||| | | | | | | | | | | | | | | | | | | | |
281 MCPXEYFDGEFTEFKNKILJVDAYN.KGNVKKEELNENETRYEIKKKP-PNY 319
      :| | | :| | | | | | | | | | | | | | | | | | | | | |
20: LSTQSALSKD-FNRKRHHNVLLSEPV 224
      :| | | :| | | | | | | | | | | | | | | | | | | | | |
3:: ---KAHNCDD.LNELDYVLKPEF 332
      :| | | :| | | | | | | | | | | | | | | | | | | | | |

RESULT 5
SY62_DISOM
ID SY62_DISOM STANDAPD PRT: 439 AA.
AC P24566;
DI U1-MAR-1992 (Rel. 2). Created;
DT 01-MAR-1992 (rel. 2), Last sequence update);
DE 01-NOV-1995 (rel. 32, last annotation updated)
GN Synaptotagmin B (Synaptic vesicle protein p65-B).
OS P65-B.
OC Discozyme omata (Electric ray);
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Teleostei; Elasmobranchii; Squales; Hypnosquales; Pristiogaster; Batelasma; Torpediniformes; Narcineidae; Narceiniidae; Discozygus.
CX NCBI_TaxId=7785;
[[[
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE
MEDLINE=91273591; PubMed=2054189;
Wondland S., Miller K.G., Seibling J., Scheller R.H.: "Differentential expression of the p65 gene family." Neuron 6:993-1007(1991).
]-] FUNCTION: MAY HAVE A REGULATORY ROLE IN THE MEMBRANE INTERACTIONS DURING TRAFFICKING OF SYNAPTIC VESICLES AT THE ACTIVE ZONE OF THE SYNAPSE. IT BINDS ACIDIC PHOSPHOLIPIDS WITH A SPECIFICITY THAT REQUIRES THE PRESENCE OF BOTH AN ACIDIC HEAD GROUP AND A DIACYL BACKBONE.
]- SUBUNIT: HOMODIMER OR HOMOTRIMER (POSSIBLE).
]- SUBCELLULAR LOCATION: SYNAPTIC VESICLES IN NEURONS.
]- TISSUE SPECIFICITY: SPINAL CORD, BRAINSTEM, MIDBRAIN AND ELECTRIC ORGAN.
]- SIMILARITY: CONTAINS 2 C2 DOMAINS.
]- SIMILARITY: BELONGS TO THE SYNAPTOTAGMIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/about/copyright/send_an_email_to_license@sib-sib.ch).

EMBS; M64276; AAA49228.1 ;
PIR: JHU0414; JHC414.
HSSP: P217C7; IBYN.
InterPro: IPRO00308; C2.
InterPro: IPRO02149; LR;.
InterPro: IPRO01365; Synaptotagmin.; Pfam: PF00166; C2_2. PRINTS: PR00340; C2DOMAIN. PROSITE: PR00399; SYNAPTOTAGMK. SMART: SM00239; C2; 2. PROSITE: PS00499; C2_DOMAIN_1; 2. PROSITE: PS00504; C2_DOMAIN_2; 2. Transmembrane; Repeat; Synapsis; Multigene family; Glycophorin. DOMAIN 1 74 VESICULAR (POTENTIAL). TRANSMEM 75 101 POTENTIAL. CYTOPLASMIC (POTENTIAL). FT DOMAIN 102 425 CYTOPLASMIC (POTENTIAL). FT DOMAIN 153 399 PHOSPHOLIPID BINDING (PROBABLE). FT DOMAIN 173 262 C2 DOMAIN 1. FT DOMAIN 304 395 C2 DOMAIN 2. FT CARBOHYD 6 6 N-LINKED GLYCANA. FT CARBOHYD 46 45 N-LINKED GLYCANA. FT CARBOHYD 439 AA: 49278 MW: 203305 DALTON 69% IDENTITY.
SEQUENCE

```


415 YLRD0058QRYWGAFIPMLTLENGDWFA... --PMEDLPFILPELVYVNGYKSP1 457

208 SKDPN 212

468 NADPN 472

RESULT 8

IRMA_CAMLE

ID TMA_CAMLE

AC Q9P92

STANDARD: PRT: 457 AA.

15-JUN-2002 (Rel. 41, Created)

15-JUN-2002 (Rel. 41, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)

trNA (Uracil-5'-methyltransferase (EC 2.1.1.35) (trNA(M-5-U54))-

methyltransferase) (RMT))

trNA OR C00851C

Campylobacter jejuni.

Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;

Campylobacter.

NCBI_TaxID=297;

[1]

SEQUENCE FROM N.A.

STRAN-MCTC 112547

MEDLINE=20150912; PubMed=10648204;

Parkhill J., Wren B.W., Kuchall K., Kellie J.M., Chaffter G.,

Basham D., Chillingworth T., Davies K.M., Paulsen O., Barrett T.,

Agarwal K., Karsberg A.V., Moule S., Pallen M.J., Parkhill J.,

Quail M.A., Rajadaram K.A., Rutherford K.E., van Vliet A.H.M.,

Whittam T.S., Barrall H.G.

*The genome sequence of the food-borne pathogen Campylobacter jejuni

reveals hypervariable sequences.*

Nature 403:665-669(2000).

-1- FUNCTION: Catalyzes the formation of 5-methyl-uracine at position

54 (M-5-U54) in all tRNA (5y similarity).

-1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA -> S-adenosyl-L-

homocysteine + tRNA containing thymine

-1- SIMILARITY: BELONGS TO THE RNA M5C METHYLTRANSFERASE FAMILY. TMA

SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL: AJ139076; CAB73096.1; .

InterPro: IPR000051; SAM_Hind.

InterPro: IPR001566; Tma.

PROSITE: PS01230; TMA.1.1.

PROSITE: PS01231; TMA.2; FA-SE-NFS.

Transferase; Methyltransferase; tRNA processing; complete proteome.

DOMAIN 207 213

S-ADENOSYL-METHIONINE BINDING (BY

SIMILARITY).

ACT_SITE 315 415

BY SIMILARITY.

SEQUENCE 357 AA; 42276 MW; CEC5328347025497 Q8064;

Query Match

Best Local Similarity 24.8%; Score 57; Db 1; Length 457;

Matches 30; Conservative 13; Mismatches 40; Gaps 5;

79 KRDFFSKAMPGVGVCGTFFPKDGNVKNVAKVNF...ENITLV-----NPIFLKG 127

111 111 111 111 111 111 111 111 111 111

14 EKHSFKKYEKFTTKFKLPASCKKHFNPAESFYHENTFLYVAFDSKSKNTVEY 73

128 IDPKED-----GNTLGKLYNNKSHNVVFNASKKRNHNVKFKIHNKIE 172

111 111 111 111 111 111 111 111 111 111

74 LDFADKICAFMPCLLVLRUNLKEKEL-----FOVEELTKQK...LSILLYHNKIE 175

111 111 111 111 111 111 111 111 111 111

173 D 173

173 D 173

DB 126 D 126

RESULT 9

DPOE_YEAST

ID DPOE_YEAST

AC 221951

STANDARD: PRT: 2222 AA.

01-AUG-1991 (Rel. 19, Created)

01-AUG-1991 (Rel. 19, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)

DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA

polymerase II subunit A)

GN POL2 OR DUN2 OR YNL262W OR NC825.

Saccharomyces cerevisiae (baker's yeast).

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

NCBI_TaxID=4932;

[1]

SEQUENCE FROM N.A. AND SEQUENCE OF 1214-1221.

MEDLINE=90381771; PubMed=2169349;

Morrison A., Azaki H., Clark A.B., Hamatake R.K., Sugino A.

A third essential DNA polymerase in S. cerevisiae.

Cell 62:1143-1151(1990).

[2]

SEQUENCE OF 1-2221 FROM N.A.

STRAN=9288C / FY:679;

MEDLINE=5430631; PubMed=8346425;

Sen Gupta K., Lyck R., Fleig U., Niedenthal R.K., Hagemann J.H.

*The sequence of a 24,152 bp segment from the left arm of chromosome

XIV from Saccharomyces cerevisiae between the BNI1 and the POL2

genes.*

Yeast 12:505-514(1996).

[3]

TEMPERATURE SENSITIVE MUTANTS.

MEDLINE=92264553; PubMed=1537345;

Araki H., Kopp P.A., Johnson A.L., Johnston L.H., Morrison A.

*DNA polymerase II, the probable homolog of mammalian DNA polymerase

epsilon, replicates chromosomal DNA in the yeast Saccharomyces

cerevisiae.*

EMBL: L11733-740(1992).

-1- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA

REPLICATION.

-1- CATALYTIC ACTIVITY: N-Deoxydeoxyriboside triphosphate + N-diphosphate

-> (dNMP)(N).

-1- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30

kDa, AND 29 kDa).

-1- SUBCELLULAR LOCATION: Nuclear.

-1- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE

N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY

FOR COMPLEXING SUBUNIT B AND C.

-1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:

ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR

DIFFERENT REACTIONS OF DNA SYNTHESIS.

-1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. HIGH

SIMILARITY WITH MAMMALIAN DNA POLYMERASE EPSILON.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL: M60416; AAA8871.1; .

EMBL: X92494; CAA63235.1; .

EMBL: Z71538; CAA36169.1; .

PIR: A36028; A36028

SGD: S0005206; POL2.

InterPro: IPR002064; DNA_pol_B.

PIfam: PF00136; DNA_pol_B.1.

PRINTS: PR0106; DNAPOCB.

DR

DR

DR

DR

DR

DR

DR SMART: SMO0486; POLBO: 1.
 DR PROSITE: PS00116; DNA POLYMERASE B; FA-SE-NES
 KW Transerase: DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding Zinc-finger; Nuclear protein;
 FT ZK_FING 2108 2191 POTENTIAL;
 FT VARIANT 544 544 M → I (IN POL-2.9 IS ETHANOL);
 FT VARIANT 710 710 P → S (IN POL-2.16 IS MUTANT);
 SQ SEQUENCE 2222 AA: 255665 MW: 687022234147165b 76742

Query Match 6.48; Score 46.5; DB 1; Length 2222;
 Best Local Similarity 28.23; Pred. No. 11;
 Matches 37; Conservative 14; Mismatches 40; Indels 3; Gaps 7;

QY 53 LPVWPP-FLVLTITGVGQ-ESRYTHM -----KRDPEKSAKEGVV-- QERT 98
 Db 983 LPKSPETYPTELENGAKLYSLVPCSLNRYVVKPTNRWDELKIDLNKIVETHEHTI 942
 QY 99 PFKDDGNYKTR-NEVFESDILVNR-----IELKIDFKEDGNLGHKLEYNYSR 146
 Db 943 FEVDGYKAMILPSSKEEGK-KKRYAVENEDSSLAELGFEKKEGRL--QLKNEQ 989
 QY 147 S-HNVVIMAD 255
 Db 100C SD-FKVFLEGD 1010

RESULT 10
 ID COPA_BOVIN STANDARD: PRI: 1224 AA.
 AC Q27951;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 43, Last annotation update)
 DE Cotomer alpha subunit (Alpha-coat protein) (Alpha-COP) (HEP-COP)
 DE (HEP-COP) [Contains: Xenin (Xenopus)-related peptide]; Proxemin;
 GN COPA.
 CS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 FN 11;
 RP SEQUENCE FROM N.A.
 RA Weland F.11;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: (MAR-1996) THE COTOMER IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS
 CC TO GLYCINE MOTIFS AND REVERSIBLY ASSOCIATES WITH DELTA NEMO-
 CC CLATHRIN-COATED VESICLES, WHICH FURTHER MEDICATE POLYMERIZING
 CC PROTEIN TRANSPORT FROM THE ER VIA THE GOLGI TO THE TRANS GOLGI
 CC NETWORK. COTOMER COMPLEX IS REQUIRED FOR BUILDING FROM GOLGI
 CC MEMBRANES, AND IS ESSENTIAL FOR THE RETROGRADE GLYT-10-ER
 CC TRANSPORT OF DILYSINE-TAGGED PROTEINS. IN MAMMALS, THE COTOMER
 CC CAN ONLY BE RECRUITED BY MEMBRANES ASSOCIATED TO ANP-R-OSYLATION
 CC FACTORS (ARGES), WHICH ARE SMALL GTP-BINDING PROTEINS. THE COMPLEX
 CC ALSO INFLUENCES THE GOLGI STRUCTURAL INTEGRITY, AS WELL AS THE
 CC PROCESSING, ACTIVITY, AND ENDOCYTIC RECYCLING OF ILL RECEPTORS (BY
 CC SIMILARITY).
 CC -!- FUNCTION: XENIN STIMULATES EXOCRINE PANCREATIC SECRETION. IT
 CC INHIBITS PENTACATRIN-STIMULATED SECRETION OF ACID, TO INHIBIT
 CC EXOCRINE PANCREATIC SECRETION AND TO AFFECT SMALL AND LARGE
 CC INTESTINAL MOTILITY. IN THE GUT, XENIN INTERACTS WITH THE
 CC NEUTENSIN RECEPTOR (BY SIMILARITY).
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX THAT CONSISTS OF AT LEAST THE ALPHA,
 CC BETA, BETA', GAMMA, DELTA, EPSILON AND ZETA SUBUNITS.
 CC -!- SUBCELLULAR LOCATION: THE COTOMER IS CYTOSOLIC OR POLYMERIZED
 CC ON THE CYTOPLASMIC SIDE OF THE GOLGI, AS WELL AS IN THE
 CC VESICLES/BUDS ORIGINATING FROM IT (BY SIMILARITY). XENIN IS
 CC SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP (=MANS)).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL/EMBL-EBI
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.ebi.ac.uk/announcement/>
 CC or send an email to license@ebi.ac.uk.)
 CC -----
 CC EMBL: X26768; CAA65543.1;
 CC InterPro: IPR001060; WD40;
 CC Pfam: PF00400; WD40.7;
 CC PRINTS: PR00320; GPR0FFIN5RPT.
 CC PRODOM: PR000018; WD40.5;
 CC SMART: SMO0320; WD40.5;
 CC PROSITE: PS00678; WD REPEATS_1; 2;
 CC PROSITE: PS00682; WD REPEATS_2; 6;
 CC PROSITE: PS00294; WD REPEATS_REG-ON; 1;
 CC Transport: Protein transport; Golgi stack; Endoplasmic reticulum;
 CC Membrane: Repeat; WD repeat; Hormone.
 FT PEPTIDE 1 25 XENIN.
 FT PEPTIDE 1 35 PROXEMIN.
 FT REPEAT 7 37 WD 1.
 FT REPEAT 49 79 WD 2.
 FT REPEAT 91 121 WD 3.
 FT REPEAT 133 163 WD 4.
 FT REPEAT 203 233 WD 5.
 FT REPEAT 247 277 WD 6.
 SQ SEQUENCE 1224 AA: 138358 MW: C9350BF2AC00693D CRC64;
 Query Match 6.78; Score 86; DB 1; Length 1224;
 Best Local Similarity 25.38; Pred. No. 16;
 Matches 40; Conservative 27; Mismatches 59; Indels 32; Gaps 7;

QY 90 EGYVOERTIFKDDGNKYTRAEVFKPGDTVNRIR-LKGTDFKEDGNLGHKLEYNYSR 146
 Db 632 KGYPEVALHFVKDE---KTRPSLALECGNIEIALEAAKALDDKNCKEKLGFVALLQSNHQ 688
 QY 149 NVYIMADKQNGIKVNF-----KIRHNIEDGGVQ--LADHYQONTPTGDSVLLPO 197
 Db 689 -VEMYQYQIKNFIDKLSFLYLITGNLEKJRKMKKIAE-RKMSGHYONALYLGD----- 741
 QY 198 NHYLSTQSALSKDPNEKRDHMYLEFVTAAGTIGMDE 235
 Db 742 --VSEVRVRLKNCGCKS-----LAYLFAA--THGLPE 769

RESULT 11
 ID CAN2_HUMAN STANDARD: PRI: 700 AA.
 AC P17555; Q14736;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Calpain 2, large (catalytic) subunit precursor (EC 3.4.22.17)
 DE (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)
 DE (Milimolar-calpain).
 GN CAPN2 OF CAPN2.
 OS Homo sapiens (Human)
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCBI_TaxID=9606;
 FN 11;
 RP SEQUENCE FROM N.A.
 RA MEDLINE:89166474; PubMed:2852952;
 RA Irajah S.; Aoki K.; Ohno S.; Emori Y.; Kawasaki H.; Suqihara H.;
 RA Suzuki K.;
 RT "Molecular cloning of the cDNA for the large subunit of the
 RT high-Ca2+-requiring form of human Ca2+-activated neutral proteinase";
 RI Biochemistry 27:8122-8128(1988).
 RN 121;
 RN SEQUENCE OF 1-79 FROM N.A.
 RC TISSUE-Lymph node;
 RX MEDLINE:89157947; PubMed:2519381;
 RA Hata A.; Ohno S.; Akita Y.; Suzuki K.;
 RT "Tandemly reiterated negative enhancer-like elements regulate
 RT transcription of a human gene for the large subunit of calcium-

dependent protease".
 C. Biol. Chem. 264:6404-6411(1989).
 [3]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RX MEDLINE:22105516; PubMed:10639123;
 RA Strobel S., Fernandez-Catalan C., Braun M., Haber P., Masumoto H.,
 RA Nakagawa K., Irie A., Sorimachi H., Bourenkov G., Bartunik H.,
 RA Suzuki K., Hode W.
 RT "The crystal structure of calcium-free human metalloprotease suggests an
 RT electrostatic switch mechanism for activation by calcium."
 RT Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000).
 CC -!- FUNCTION: Calcium-regulated on lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr- Xaa, Met- Xaa or
 CC Arg-1-Xaa with Leu or Val as the P2 residue.
 CC -!- COFACTOR: Binds 3 calcium ions.
 CC -!- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
 CC of calcium and inhibited by calpastatin.
 CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding.
 CC -!- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM BINDING DOMAINS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 02.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M23254; AAA35645.1; -
 DR EMBL: J04706; AAAS2760.1; -
 DR PIR: A31218; A31218.
 DR PDB: 1KFU; C7-DEP-G1.
 DR MEROPS: C02.002; -
 DR Genbank: HGNC:1475; CALN2.
 DR MIM: 114230; -
 DR InterPro: IPR002548; EF-hand.
 DR InterPro: IPR001300; Protease_C2.
 DR InterPro: IPR003169; SHP1-related.
 DR Pfam: PF00036; ef-hand; 3.
 DR Pfam: PF00648; Papain-like; 1.
 DR Pfam: PF01067; Calpain_1; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR ProDom: PD000012; EF-hand; 1.
 DR SMART: SM00230; Cyspe; 1.
 DR SMART: SM00054; EFh; 2.
 DR PROSITE: PS00018; EF-HAND; 2.
 DR PROSITE: PS00139; THIOL-PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL-PROTEASE_HIS; FALSE-NEG.
 DR PROSITE: PS00640; THIOL-PROTEASE_ASN; FALSE-NEG.
 KW Hydrolase; Thiol protease; Calcium-binding; Repeat: Multiple family;
 KW 3D-structure.
 FT PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
 FT CHAIN 20 790 CALPAIN 2, LARGE (CATALYTIC) SUBUNIT.
 FT DOMAIN 20 210 THIOL-PROTEASE DOMAIN 1.
 FT DOMAIN 211 395 THIOL-PROTEASE DOMAIN 11.
 FT DOMAIN 355 514 THIOL-PROTEASE DOMAIN 11.
 FT DOMAIN 515 529 LINKER.
 FT DOMAIN 530 700 DOMAIN IV.
 FT DOMAIN 541 552 EF-HAND 1.
 FT CA_BIND 585 596 EF-HAND 2.
 FT CA_BIND 615 626 EF-HAND 3.
 FT CA_BIND 650 661 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT DOMAIN 680 691 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT ACT_SITE 135 165 BY SIMILARITY.
 FT ACT_SITE 242 262 BY SIMILARITY.
 FT ACT_SITE 286 286 BY SIMILARITY.
 FT CONFLICT 68 S 75 G (IN REF. 2).

CU CONELICI 73 74 MR -> IE (IN REF. 2)
 SO SEQUENCE 700 AA: 80006 MR: A944D13BC8465531 CRC64;
 Query Match 6.7%; Score 85.5; DB 1; Length 700;
 Best Local Similarity 20.9%; Pred. No. 9.2;
 Matches 47; Conservative 39; Mismatches 80; Indels 59; Gaps 10;
 QY 4 GBEIET---GVVPIVLELDGVNHHK-----FSVSGEGGDATYKLTILKFKICT--GKLPVP 55
 DB 423 GDMHTITGFIYVEPEELSGQTNHLSKNFFLNNHAKERSDTF--INLREVNRFKLP-P 479
 QY 57 WPTLVITLYG-----VQCFS-RYEDHMKRHFPSKAMPEGVVQERTFFEDDGNVX 107
 DB 480 GSYIVPSTFFEPKKGDFCIRVFSEKKADYQAVDDIEANLEEDFISDDI---QDAVRR 536
 QY 106 TRAFV-----KFEGETLVNRIELKQIDFKNDG----- 134
 DB 517 LPAQLAGSAEIASAPELOILFRVLAKRQDIKDSFETICKIMVDM:JNSGSKLGKE 596
 QY 135 -NLQHKLFYNNSHNYIMADKKNG:KVNFKIHUNIED:GVOL 178
 DB 507 FVILWTKG---KYQKIVSEIDVDSGTNSYEMKKALEHAGPKM 639
 RESULT 12
 D152_HABIN STANDARD PRI: 795 AA.
 AC P44955;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DE (D-15-Ag) (Outer membrane protein D15).
 GN H109.7
 OS Haemophilus influenzae
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales
 OC Haemophilus
 OX NCBI_TaxID:727;
 RN 1;
 RP SEQUENCE FROM N.A.
 RC STRAIN=BJ / KW20 / ATCC 51907;
 RX MEDLINE:95350630; PubMed:7542860;
 RA Fleischmann R.D., Adams M.D., White G., Clayton P.A., Kirkness E.F.,
 RA Kiehlavage A.R., Sult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Min I.-L., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.H.,
 RA Jeterback I.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.S., Fritchman J.L., Fuhrmann J.L., Goehagen N.S.M.,
 RA Greth C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RA Rd."
 RL Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U32773; AAC22575.1; -
 DR TIGR: H10517; -
 DR InterPro: IPR000184; Bac_surfAq_D15.
 DR Pfam: PF01103; Bac_surfAq; 1.
 DR Antigen: Outer membrane; Signal; Complete proteome.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
 FT SEQUENCE 795 AA: 87478 MR: B85691FC22E6ED44 CRC64;

Query Match 6.78, Score 45.5, DP 11, Length 797;
 Best Local Similarity 22.68, Pred. No. 11;
 Matches 49: Conservative 33; Mismatches 72; Indels 43; Gaps 12;

66 YGVQCFSHYPDHMKHDP-----PKSAMPEYVQE-----RTIPK 170
 429 VGTESG:SYQASVQGNFLG:GAAVS:ANTANDYGTNVN:GAYTPYETK:GVSL:GVNFF 488
 101 KDGNYKTRAEVK:PGDTLVKRIE:KGDFKEDNI---LQK-----KLVNYSNHN 149
 483 ENYNSKSCSTSNYKRTYGSNVTL:GPPVWKNNSYVYGLGHTYKNKISNFALEYN---RN 544
 150 VYIMADKKK-NELKVN-----FKLRHN-----LEDGGVQLADHYQNTIL:GSGVILP- 195
 545 LVIQSMKTKGNSIKTNDGDFSGNWNINSLNGYPTTKGVKAS-----LS-GRVTIG 595
 197 -DNHYLSTQSA:SKDPNEK:RHMVLEFVTAAGLHG 232
 596 SGNKYKLSADVQGGYPLDRDHLWVWSAKASAGYANG 632

RESULT 13

D151_HAEIN
 ID D151_HAEIN STANDARD: PRT: 797 AA.
 AC P46024;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 43, Last annotation update)
 DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DE (D-15-Ag) Outer membrane protein D15.
 DE Haemophilus influenzae.
 OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
 OC Haemophilus
 OC Haemophilus
 CX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Serotype B;
 RX MEDLINE=92555676; PubMed=7717523;
 RA Black P.S., Loosmore S., Chong P., Thomas W.P.
 SC "The sequencing of the 80-KDa D15 protective surface antigen of
 PT Haemophilus influenzae."
 PJ Gene 136:97-99(1995).
 PL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Minda, and Egan / Serotype B;
 RX MEDLINE=97427052; PubMed=9264146;
 RA Loosmore S.M., Yang Y.P., Coleman D.S., Shortness J.M., and Land D.M.,
 Klein M.H.;
 SC "Outer membrane protein D15 is conserved among Haemophilus influenzae
 PT species and may represent a universal protective antigen against
 KT invasive disease."
 RL Infect. Immun. 65:3701-3707(1997).
 CC -1- SURCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/usage/new/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U13961; AA085645.1;
 CC DR U60832; AAB61974.1;
 CC EMBL: U60833; AAB61976.1;
 CC InterPro: IPR000184; IAC_SurTAO_D15.
 CC Pfam: PF01103; Bac_SurTAO_Ag; 1;
 CC Antigen: Outer membrane; Signal;
 CC SIGNAL 20 797 POTENTIAL,
 CC CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
 CC SEQUENCE 797 AA: 87675 MW: 2F93DE538696A1B CRC54;

Query Match 6.78, Score 45.5, DP 11, Length 797;
 Best Local Similarity 22.68, Pred. No. 11;
 Matches 49: Conservative 33; Mismatches 72; Indels 43; Gaps 12;

66 YGVQCFSHYPDHMKHDP-----PKSAMPEYVQE-----RTIPK 170
 429 VGTESG:SYQASVQGNFLG:GAAVS:ANTANDYGTNVN:GAYTPYETK:GVSL:GVNFF 488
 101 KDGNYKTRAEVK:PGDTLVKRIE:KGDFKEDNI---LQK-----KLVNYSNHN 149
 483 ENYNSKSCSTSNYKRTYGSNVTL:GPPVWKNNSYVYGLGHTYKNKISNFALEYN---RN 544
 150 VYIMADKKK-NELKVN-----FKLRHN-----LEDGGVQLADHYQNTIL:GSGVILP- 195
 545 LVIQSMKTKGNSIKTNDGDFSGNWNINSLNGYPTTKGVKAS-----LS-GRVTIG 595
 197 -DNHYLSTQSA:SKDPNEK:RHMVLEFVTAAGLHG 232
 596 SGNKYKLSADVQGGYPLDRDHLWVWSAKASAGYANG 632

RESULT 14

CAN2_PIG
 ID CAN2_PIG STANDARD: PRT: 324 AA.
 AC P43367;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Calpain 2, large (catalytic) subunit (EC 3.4.22.17) (Calcium-activated
 DE neutral proteinase) (CAMP) (M-type) (M-calpain) (Millimolar-calpain)
 DE (Fragment).
 DE CAPN2.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 CX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE OF 1-209 FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=94146155; PubMed=812396;
 RA Sun W., Ji S.Q., Ebert R.J., Bidwell C.A., Hancock J.B.;
 SC "Cloning the partial cDNAs of mu-calpain and m-calpain from porcine
 PT skeletal muscle."
 PL Biochimie 75:931-936(1993).
 RN [2]
 RP SEQUENCE OF 122-324 FROM N.A.
 RC TISSUE=pulmonary artery;
 RX MEDLINE=98400921; PubMed=9728340;
 RA Zhang J.L., Patel J.M., Block F.R.;
 SC "Hypoxia-specific upregulation of calpain activity and gene
 PT expression in pulmonary artery endothelial cells."
 RL Am. J. Physiol. 275:H461-H468(1998).
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodelling and signal transduction (By similarity).
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the P2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions.
 CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
 CC of calcium and inhibited by calpastatin.
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding.
 CC -1- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/usage/new/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U13961; AA085645.1;
 CC DR U60832; AAB61974.1;
 CC EMBL: U60833; AAB61976.1;
 CC InterPro: IPR000184; IAC_SurTAO_D15.
 CC Pfam: PF01103; Bac_SurTAO_Ag; 1;
 CC Antigen: Outer membrane; Signal;
 CC SIGNAL 20 797 POTENTIAL,
 CC CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
 CC SEQUENCE 797 AA: 87675 MW: 2F93DE538696A1B CRC54;

CC entities requires a license agreement (See <http://www.list-sip.ch/announce/> or send an email to license@list-sip.ch).

CC	EMBL: U01182; AAC48401; ...				
DR	EMBL: U71320; AAB17391; ...				
DR	HSSP: Q07009; IDFO.				
DR	MEROPS: CC2.002; ...				
DR	InterPro: IPR002048; EF-hand.				
DR	InterPro: IPR001300; Protease C2.				
DR	InterPro: IPR000159; SH3COLacsite.				
DR	Pfam: PF00336; eihand; 2.				
DR	Pfam: PF01067; Catpala1; 1.				
DR	PRODOM: PD060012; EF-hand; 1.				
DR	PROSITE: PS00018; EF-hand; 1.				
DR	PROSITE: PS00139; THIOL PROTEASE CYS; PARTIAL.				
DR	PROSITE: PS00639; THIOL PROTEASE HIS; PARTIAL.				
DR	PROSITE: PS00643; THIOL PROTEASE ASN; PARTIAL.				
KW	Hydrolase; Thiol protease; Calcium-binding; Repeat; Multigene family.				
FT	NOTER	1			
FT	DOMAIN	157			
FT	DOMAIN	158	324		
FT	DOMAIN	138			
FT	DOMAIN	139			
FT	DOMAIN	154	324		
FT	CA_BIND	165	176		
FT	CA_BIND	209	220		
FT	CA_BIND	239	250		
FT	DOMAIN	274	285		
FT	DOMAIN	304	315		
FT	CONFUCT	202	202		
FT	SEQUENCE	324 AA;	37808 MW;	35235532-36123CF	36742

[illegible]

RESULT	ID	STANDARD	PR	PR
D153_HAEIN	D153_HAEIN	753 AA		
AC	Q32629			
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Protective surface antigen D15 precursor (80 kDa D15 antigen)			
DE	(0-15-Ag) Outer membrane protein (15)			
OS	Haemophilus influenzae			
CC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellales			
CC	Haemophilus			
CC	NCBI_TaxID:727			
KN	[...]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-PAK 12085			

MEDLINE:37427952; PubMed-9294140;
 Loomsme R.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
 Klein M.H.;
 "Outer membrane protein D15 is conserved among Haemophilus influenzae
 species and may represent a universal protective antigen against
 invasive disease";
 Infect. Immun. 65:3701-3707 (1997).
 -1- SUBCELLULAR LOCATION: Outer membrane.
 -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.ebi.ac.uk/submit/>
 or send an email to license@ebi.ac.uk).

```

EMBL: U60834: AAR61977.1
InterPro: IPR00184: Bact_Sulfag_D15.
Flam: PF0103: Bact_sulfag_M9: 1
Antigen: Outer membrane; Signal.
SIGNAL 1 19 POTENTIAL.
CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SEQUENCE 793 AA: 8755 MW: 518FDH20486C1A14 CRC64:

Query Match
Best Local Similarity 23 CA: Pred No. 14:
Matches 50: Conservative 31: Mismatches 73: Indels 63: Gaps 12:

56 YGVQCFSPYDHNKRF-----FKSAMPEGVGE-----R:IFF 100
||| : : : : : |||
429 YGTSGISVQTSKQDNEVGTCGAVSAGIKNDYGTSVNLGYTEPYTKQVSLGSGIFF 488
||| : : : : : |||
102 KQDGNKYTKAAYKEGDTLVNTEKLGIDFKDGN-----LGH-----KLFVNYNSHN 149
||| : : : : : |||
489 ENYDNSKSTSSNYKATYGSNVLGEPVNNNSVYVGLGHYIKNSIALEYN-----RN 544
||| : : : : : |||
150 VYINAKQKNGKVN-----FKRIIN-----IDGGVQLADHYQNTPLDEGPVLLP- 196
||| : : : : : |||
545 LYIQSKFKGKGIKKEPSESGWYNSLNROYFTKGVKAS-----LGGRVTIG 635
||| : : : : : |||
197 -DNHYLSIGASLKIDNFKRDNKVLLEVTANGTIG 242
||| : : : : : |||
596 SDNKYKLSAVQGPYPLDRDHRWYVSAKASAGYANG 632
||| : : : : : |||

```

```
Search completed: July 26, 2003, 16:11:19
Job time : 11.6667 secs
```

GeneCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein - protein search, using sw model

Run on: July 16, 2003, 16:06:07 ; Search time 41.33s Seconds
(without alignments)
1186,423 Million cell updates/sec

Title: us-09-967-301-4
Perfect score: 1276
Sequence: 1 MSKSGCFGVVPIVELDQ.....VILFFVTAAGTTCGMLYK AAF

Scoring table: RUCSOM62
Gapop 10.0 ; Gapext 6.5

Searched: 671360 seqs, 206247115 residues

Total number of hits satisfying chosen parameters: 671360

Minimum DB seq length: 6
Maximum DB seq length: 200000000

Post-processing: Minimum Match: 04
Maximum Match: 204
Listing: 1 first 45 summaries

Database:

SPIREML_21:

- 1: sp-archaei*
- 2: sp-bacteria*
- 3: sp-fungi*
- 4: sp-human*
- 5: sp-invertebrates*
- 6: sp-mammal*
- 7: sp-misc*
- 8: sp-ornithelie*
- 9: sp-plauro*
- 10: sp-plant*
- 11: sp-rodent*
- 12: sp-virus*
- 13: sp-vertebrate*
- 14: sp-unclassified*
- 15: sp-virus*
- 16: sp-bacteriap*
- 17: sp-archaeap*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:

Result NO.	Score	Query Match	Length	DB ID	Description
1	1258	98.6	238	5 Q27953	Q27953 unidentified
2	1252	98.1	238	5 Q91125	Q91125 aequorea vi
3	1216	95.3	238	5 Q17105	Q17105 aequorea vi
4	1215	95.2	758	4 Q96001	Q96001 hmo sapien
5	1201	94.1	238	5 Q17106	Q17106 aequorea vi
6	1109	86.9	238	5 Q8WTC6	Q8WTC6 aequorea ma
7	1105	86.6	238	5 Q8WPT5	Q8WPT5 aequorea ma
8	1102	86.4	238	5 Q8WTC4	Q8WTC4 aequorea ma
9	1100	86.2	238	5 Q8WTC0	Q8WTC0 aequorea ma
10	1099	86.1	238	5 Q8WTC5	Q8WTC5 aequorea ma
11	1099	86.1	238	5 Q8WTC8	Q8WTC8 aequorea ma
12	1096	85.9	238	5 Q8WTC7	Q8WTC7 aequorea ma
13	1095	85.8	238	5 Q8WTC3	Q8WTC3 aequorea ma
14	251.5	19.7	225	5 Q950A7	Q950A7 montastraea
15	243.5	19.6	236	5 Q8T606	Q8T606 dendronephr
16	249	19.5	225	5 Q953F5	Q953F5 montastraea

17	243.5	19.1	266	5 Q9U6Y3	Q9U6Y3 clavularia
18	236.5	16.5	225	5 Q8T5F1	Q8T5F1 montastraea
19	210.5	16.5	234	5 Q8T5F2	Q8T5F2 montastraea
20	210	16.5	229	5 Q9U6Y6	Q9U6Y6 anemonia ma
21	209.5	16.4	225	5 Q9J5V9	Q9J5V9 discosoria s
22	209	16.4	221	5 Q95P04	Q95P04 conioptora t
23	208	16.3	227	5 Q962P9	Q962P9 montastraea
24	205.5	16.1	232	5 Q9GPT5	Q9GPT5 anemonia su
25	204.5	16.0	235	5 Q8T5F0	Q8T5F0 scolymia cu
26	204	16.0	227	5 Q95V70	Q95V70 montastraea
27	204	16.0	238	5 Q9BLV9	Q9BLV9 renilla mue
28	203	15.9	232	5 Q9U6Y7	Q9U6Y7 discosoma s
29	201.5	15.8	232	5 Q93Z28	Q93Z28 anemonia su
30	197.5	15.3	225	5 Q8T6T9	Q8T6T9 heteractis
31	195	15.3	230	5 Q9G137	Q9G137 discosoma s
32	187.5	14.7	227	5 Q95W85	Q95W85 heteractis
33	187	14.7	228	5 Q9GB16	Q9GB16 anemonia su
34	182.5	14.3	231	5 Q9U6Y5	Q9U6Y5 zoanthus sp
35	181.5	14.2	238	5 Q9BL2C	Q9BL2C plilosarcus
36	181	14.2	233	5 Q96319	Q96319 renilla ren
37	179.5	14.1	227	5 Q95W96	Q95W96 condylactis
38	175	13.7	229	5 Q8T5B7	Q8T5B7 condylactis
39	173.5	13.6	227	5 Q95W11	Q95W11 condylactis
40	173.5	13.6	241	5 Q8T5B8	Q8T5B8 ridoidea fl
41	171	13.4	231	5 Q9U6Y4	Q9U6Y4 zoanthus sp
42	166.5	13.0	231	5 Q8T4M4	Q8T4M4 zoanthus sp
43	165	12.9	234	5 Q8T5F3	Q8T5F3 scolymia cu
44	152.5	12.0	241	5 Q8T6T8	Q8T6T8 discosoma s
45	142	11.1	231	5 Q8T5E9	Q8T5E9 ridoidea fl

ALIGNMENTS

RESULT:

Q27953 PRELIMINARY: PR: 238 AA.
AC Q27953
CT 01-NOV-1996 (TRENBLER, C.: Created)
LI 01-NOV-1996 (TRENBLER, C.: Last sequence update)
RT 01-JUN-2002 (TRENBLER, C.: Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID:12644:
RN 11
R1 SEQUENCE FROM N.A.
R2 MEDLINE:97259832; PubMed:9154981;
R3 Kowandall G.J.A., Mendes G., Wolbert E.J.H., de Boer A.B.
R4 "Enhanced expression in tobacco of the gene encoding green fluorescent
R5 protein by modification of its codon usage."
R6 Plant Mol Biol 43:889-900(1997).
R7 EMSL: X96418; CAA65278.1;
R8 HSP: P42212; IGFL.
R9 InterPro: IPR00786; Green_fl_protein.
R0 Pfam: P50353; GFP.
R1 PRINTS: PR01229; GFP-UNDESCENT.
R2 ProDom: PD013756; Green_fl_protein; 1.
R3 SEQUENCE: 238 AA; 25884 MW; CA932D47262AF2D3 CRC64;

Query Match 98.6%, Score 1258, DB 5, Length 238
Best Local Similarity 98.3%, Pred. No. 8.2e-98
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY	1	MSKRELPFGVPIVELDQVNVNKHKFSVSGEGSDATYKGLTKLFCTTGKLPVWPVL	60
DE	1	MKGRELPFGVPIVELDQVNVNKHKFSVSGEGSDATYKGLTKLFCTTGKLPVWPVL	60
QY	61	VTITLVGVCFSKYPDHMKRHDFEKSAMPGYVCGERTIFPKDGNKYNKTRAEVKEGDTLV	120
DE	61	VTITLVGVCFSKYPDHMKRHDFEKSAMPGYVCGERTIFPKDGNKYNKTRAEVKEGDTLV	120


```

SQ SEQUENCE 758 AA: 85015 MW: 8612408660707049 CRC64:
Query Match 95.2%; Score 12.5; DB 5; Length 258;
Best Local Similarity 95.6%; Pred. No. 1.5e-95;
Matches 228; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDQATYGLKLTFCITGKLPVWPTL 60
DB 499 VSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDQATYGLKLTFCITGKLPVWPTL 558
QY 61 VTITLYGVQCFSPYDPMKQKHDFKSAPEGYVOERTIFFKDGKYNKTAENKVFECTIV 120
DB 499 VTITLYGVQCFSPYDPMKQKHDFKSAPEGYVOERTIFFKDGKYNKTAENKVFECTIV 612
QY 121 NRLEKLGIDFKEDGNLGHKLEYNYSNHNYYIMADKCKNKTKVNEKIRHNIESSGVQLAD 180
DB 499 NRLEKLGIDFKEDGNLGHKLEYNYSNHNYYIMADKCKNKTKVNEKIRHNIESSGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNKSKRDHMLVLEFVTAAGTTHGMDELK 238
DB 499 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNKSKRDHMLVLEFVTAAGTTHGMDELK 238
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNKSKRDHMLVLEFVTAAGTTHGMDELK 238
DB 499 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNKSKRDHMLVLEFVTAAGTTHGMDELK 238

RESULT 5
Q17106 PRELIMINARY: PRT: 238 AA.
AC Q17106
QY 01-NOV-1996 (TrEMBLrel. 01, Created)
DB 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN 1
RP SEQUENCE FROM N.A.
RA Watkins J.A., Campbell A.K.,
RC Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: X83960; CAA58790.1;
DK HSSP: P42222; IRFP.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP.
DR PRINTS: PR01229; GFPLORESCENT.
DR PRODOM: PD013756; Green_fl_protein; 1.
FT NON_T58 238
SQ SEQUENCE 238 AA: 25867 MW: 804646262084804 CRC64:

Query Match 94.1%; Score 12.0; DB 5; Length 238;
Best Local Similarity 93.3%; Pred. No. 5e-93;
Matches 222; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDQATYGLKLTFCITGKLPVWPTL 60
DB 499 VSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDQATYGLKLTFCITGKLPVWPTL 60
QY 61 VTITLYGVQCFSPYDPMKQKHDFKSAPEGYVOERTIFFKDGKYNKTAENKVFECTIV 120
DB 499 VTITLYGVQCFSPYDPMKQKHDFKSAPEGYVOERTIFFKDGKYNKTAENKVFECTIV 120
QY 121 NRLEKLGIDFKEDGNLGHKLEYNYSNHNYYIMADKCKNKTKVNEKIRHNIESSGVQLAD 180
DB 499 NRLEKLGIDFKEDGNLGHKLEYNYSNHNYYIMADKCKNKTKVNEKIRHNIESSGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNKSKRDHMLVLEFVTAAGTTHGMDELK 238
DB 499 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNKSKRDHMLVLEFVTAAGTTHGMDELK 238

RESULT 6
Q8WTC6 PRELIMINARY: PRT: 238 AA.
AC Q8WTC6
QY 01-MAR-2002 (TrEMBLrel. 20, Created)
DB 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN 1
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RC Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY013821; AAK02059.1;
DK InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27049 MW: 81850055290128 CRC64:

Query Match 86.6%; Score 11.05; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 5.6e-95;

```

```

AC Q8WTC6
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN 1
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RC Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435431; AA53316.1;
DK InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27015 MW: 658107588926903 CRC64:

Query Match 86.3%; Score 11.59; DB 5; Length 238;
Best Local Similarity 83.8%; Pred. No. 2.6e-95;
Matches 199; Conservative 20; Mismatches 19; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDQATYGLKLTFCITGKLPVWPTL 60
DB 499 VSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDQATYGLKLTFCITGKLPVWPTL 60
QY 61 VTITLYGVQCFSPYDPMKQKHDFKSAPEGYVOERTIFFKDGKYNKTAENKVFECTIV 120
DB 499 VTITLYGVQCFSPYDPMKQKHDFKSAPEGYVOERTIFFKDGKYNKTAENKVFECTIV 120
QY 121 NRLEKLGIDFKEDGNLGHKLEYNYSNHNYYIMADKCKNKTKVNEKIRHNIESSGVQLAD 180
DB 499 NRLEKLGIDFKEDGNLGHKLEYNYSNHNYYIMADKCKNKTKVNEKIRHNIESSGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNKSKRDHMLVLEFVTAAGTTHGMDELK 238
DB 499 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNKSKRDHMLVLEFVTAAGTTHGMDELK 238

RESULT 7
Q8WTC6 PRELIMINARY: PRT: 238 AA.
AC Q8WTC6
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN 1
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qiu Y.X., Pang S.Q.,
RC Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY013821; AAK02059.1;
DK InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27049 MW: 81850055290128 CRC64:

Query Match 86.6%; Score 11.05; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 5.6e-95;

```

```

Matches 198: Conservative 20, Mismatches 20, Indels 0, Gaps 0
QY 1 MSKGEELFTGVVPLVELDGVNGHGFVSQGESEATYKGLKFKCTITTKLPPWPTL 60
DB 1 MSKGEELFTGVVPLVELDGVNGHGFVSQGESEATYKGLKFKCTITTKLPPWPTL 60
QY 61 VTTLYGVQCFSPYPMHKKHDFPKSAMPEGYVQERTIFFKDGNYKTRAEVKFESDTLV 120
DB 61 VTTLYGVQCFSPYPMHKKHDFPKSAMPEGYVQERTIFFKDGNYKTRAEVKFESDTLV 120
QY 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKVNFKIRHNIEDGGVOLAD 180
DB 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKVNFKIRHNIEDGGVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSQSKDPNEKRDMVLEFVTAAGTTHGMDLYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTALSQSKDPNEKRDMVLEFVTAAGTTHGMDLYK 238

RESULT 8
Q8WTC4 PRELIMINARY: PRT: 238 AA.
ID Q8WTC4: 86.44; Score 1102; DB 5; Length 238;
AC Q8WTC4: 84.5%; Pred. No. 10-84;
Matches 201: Conservative 16, Mismatches 20, Indels 0, Gaps 0;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DI 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435427; AAL33913.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF76E5A73A CRC64;

Query Match 86.44; Score 1102; DB 5; Length 238;
Best Local Similarity 84.5%; Pred. No. 10-84;
Matches 201: Conservative 16, Mismatches 20, Indels 0, Gaps 0;
QY 1 MSKGEELFTGVVPLVELDGVNGHGFVSQGESEATYKGLKFKCTITTKLPPWPTL 60
DB 1 MSKGEELFTGVVPLVELDGVNGHGFVSQGESEATYKGLKFKCTITTKLPPWPTL 60
QY 61 VTTLYGVQCFSPYPMHKKHDFPKSAMPEGYVQERTIFFKDGNYKTRAEVKFESDTLV 120
DB 61 VTTLYGVQCFSPYPMHKKHDFPKSAMPEGYVQERTIFFKDGNYKTRAEVKFESDTLV 120
QY 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKVNFKIRHNIEDGGVOLAD 180
DB 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKVNFKIRHNIEDGGVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSQSKDPNEKRDMVLEFVTAAGTTHGMDLYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTALSQSKDPNEKRDMVLEFVTAAGTTHGMDLYK 238

RESULT 9
Q8WTC4 PRELIMINARY: PRT: 238 AA.
ID Q8WTC4: 86.44; Score 1099; DB 5; Length 238;
AC Q8WTC4: 83.2%; Pred. No. 1.8e-84;
Matches 198: Conservative 19, Mismatches 21, Indels 0, Gaps 0;
DT 01-MAR-2002 (Tremblrel. 21, Last annotation update)
DI 01-MAR-2002 (Tremblrel. 21, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435428; AAL33913.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A192174CP64D CRC64;

Query Match 86.1%; Score 1099; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 1.8e-84;
Matches 198: Conservative 19, Mismatches 21, Indels 0, Gaps 0;

```

```

LT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435427; AAL33912.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192174CP64D CRC64;

Query Match 86.2%; Score 1100; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 1.5e-84;
Matches 198: Conservative 19, Mismatches 21, Indels 0, Gaps 0;
QY 1 MSKGEELFTGVVPLVELDGVNGHGFVSQGESEATYKGLKFKCTITTKLPPWPTL 60
DB 1 MSKGEELFTGVVPLVELDGVNGHGFVSQGESEATYKGLKFKCTITTKLPPWPTL 60
QY 61 VTTLYGVQCFSPYPMHKKHDFPKSAMPEGYVQERTIFFKDGNYKTRAEVKFESDTLV 120
DB 61 VTTLYGVQCFSPYPMHKKHDFPKSAMPEGYVQERTIFFKDGNYKTRAEVKFESDTLV 120
QY 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKVNFKIRHNIEDGGVOLAD 180
DB 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKVNFKIRHNIEDGGVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSQSKDPNEKRDMVLEFVTAAGTTHGMDLYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTALSQSKDPNEKRDMVLEFVTAAGTTHGMDLYK 238

RESULT 10
Q8WTC5 PRELIMINARY: PRT: 238 AA.
ID Q8WTC5: 86.44; Score 1102; DB 5; Length 238;
AC Q8WTC5: 84.5%; Pred. No. 10-84;
Matches 201: Conservative 16, Mismatches 20, Indels 0, Gaps 0;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DI 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435428; AAL33913.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A192174CP64D CRC64;

Query Match 86.1%; Score 1099; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 1.8e-84;
Matches 198: Conservative 19, Mismatches 21, Indels 0, Gaps 0;

```

QY	1	MSKGEELFTGVVPIILVELDGVNGHKFSVSGEAGEGADATYAKLFLAKICTIKLIPVWPTL	60
DB	1	MSKGEELFTGVVPIILVELDGVNGHKFSVSGEAGEGADATYAKLFLAKICTIKLIPVWPTL	60
QY	61	VTLITGVQCFSRYPDHNKRIHCFKSAMPEGYVGEKTIPEKTDGNTKTKTAENKFFESDLY	120
DB	61	VTLITGVQCFSRYPDHNKRIHCFKSAMPEGYVGEKTIPEKTDGNTKTKTAENKFFESDLY	120
QY	121	NRIELKGIDFKEDGNLGHKSLEYYNHSNIVYIMADKKIKYVKFIRHNIIDGGVQLAL	180
DB	121	NRIELKGIDFKEDGNLGHKSLEYYNHSNIVYIMADKKIKYVKFIRHNIIDGGVQLAL	180
QY	181	HYQNTPTIGDSVPLIPDHNHLYSTLSALSKDNKRDHNVLLERVTAAATGHGMHLYK	240
DB	181	HYQNTPTIGDSVPLIPDHNHLYSTLSALSKDNKRDHNVLLERVTAAATGHGMHLYK	240

RESULT 12

ID	Q8WTC3	PRELIMINARY:	PRT:	238 AA:
AC	Q8WTC3:			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Acquired macrodactyla.			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Hydrozoa; Hydrozoa; Leptomedusae;			
CC	Acquired macrodactyla.			
CX	NCBI:taxid=147619;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=GFPXN163;			
RA	Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan R.Q., Chen X.			
RT	"Colocal mutants of green fluorescent protein from Anopheles macrodactyla."			
RI	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.			
RF	EMBL: AF435429; AAL33914.1;			
DR	InterPro: IPR000786; Green fluorescent protein.			
DR	PIfam: PF01353; GFP.1.			
DR	Prodom: P0013756; Green fluorescent protein.			
SQ	SEQUENCE 238 AA: 27037 MW: 555418 DALY 1843 RES:			

Query Match: 26.1%; Score 1099; 66 % Ident. 238

Best local similarity: 23.2%; Evid. No. 1,26-84;

Matches 198; Conservative 19; Mismatches 21; Indels 1; Gaps 0

QY	1	MSKGEELFTGVVPIILVELDGVNGHKFSVSGEAGEGADATYAKLFLAKICTIKLIPVWPTL	60
DB	1	MSKGEELFTGVVPIILVELDGVNGHKFSVSGEAGEGADATYAKLFLAKICTIKLIPVWPTL	60
QY	61	VTLITGVQCFSRYPDHNKRIHCFKSAMPEGYVGEKTIPEKTDGNTKTKTAENKFFESDLY	120
DB	61	VTLITGVQCFSRYPDHNKRIHCFKSAMPEGYVGEKTIPEKTDGNTKTKTAENKFFESDLY	120
QY	121	NRIELKGIDFKEDGNLGHKSLEYYNHSNIVYIMADKKIKYVKFIRHNIIDGGVQLAL	180
DB	121	NRIELKGIDFKEDGNLGHKSLEYYNHSNIVYIMADKKIKYVKFIRHNIIDGGVQLAL	180
QY	181	HYQNTPTIGDSVPLIPDHNHLYSTLSALSKDNKRDHNVLLERVTAAATGHGMHLYK	240
DB	181	HYQNTPTIGDSVPLIPDHNHLYSTLSALSKDNKRDHNVLLERVTAAATGHGMHLYK	240

RESULT 12

ID	Q8WTC7	PRELIMINARY:	PRT:	238 AA:
AC	Q8WTC7:			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Green fluorescent protein.			

```

DB 1 MSKRELEFGVYVILVELGQVHGKESVROGEGQADYVQKLEIRFICITISKLVIMPTL 60
QY 61 VTLLYGVQCSRYDHRKRDFFKSAMPEYVJRTIFFKQDQNYKTRASVREJILY 120
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 62 VTLLGGLGAFARFHHKNNKFFKSAPRYTGFRTIFFKQDQNYKTRASVREJILY 120
QY 121 NRIELGIDFKHGNKTKGRLEFYNNNSHNVYIMAKKQKSKVNFKIDNLEQNVLAG 180
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 121 NRIELKGMDFKEDSNLJARKLEYNPNNSHNVIMFFKANKGIKVNKCKIDNLEQNVLAG 180
QY 181 HYQQTPTGLGQVLPENNYLSTQSAQSKDNEKHEKVEVLEFYAALINIMDELK 240
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 181 HYQTNVPLGQVLPENNYLSTQSAQSKDNEKHEKVEVLEFYAALINIMDELK 240

```

RESULT 14

```

Q95DA7 PRELIMINARY: PRT: 225 AA.
AC Q95DA7
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DI 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cyan fluorescent protein (fragment).
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Scleractinia;
QC Pavlovina; Pavlovidae; Montastraea.
GX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Falkowski P.G., Sun Y.
RT "Montastraea cavernosa fluorescent protein."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases
DR EMBL: AF056460; AAL17965.1;
DR InterPro: IPR000786; Green_fli_protein.
DR Pfam: PF01553; GFP: 1.
DR ProDom: PD013756; Green_fli_protein: 1.
FT NON_TER 225
FT TER 225
SQ SEQUENCE 225 AA: 25775 MW: 52DE2F716084524 CMC64.

```

```

Query Match 19.7%; Score 251.5; DB 5; Length 225;
Best Local Similarity 28.3%; Pred. No. 2.2e-13;
Matches 67; Conservative 48; Mismatches 81; Indels 41; Gaps 9;

QY 11 VVPLVELGQVHGKESVROGEGQADYVQKLEIRFICITISKLVIMPTL 60
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 2 VHK:KLWKGIVNGHKEMITGSEKPEPTGILIKYKESGPIVWGLTATAPVGNK 60
QY 70 GSRPDPDKRIDTFKSAMPEYVQERTIFFKQDQNYKTRASVREJILY 120
DB 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 67 VETKPKDIFDYKQSPPEYNSWERSKREAGVTVTSIKLEIDGTFYETKRYGVN 124
QY 130 FKEDGNILGKHKLEYNNNSHNVYIMAKKQKSKVNFKIDNLEQNVLAG 180
DB 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 125 FTSSGVPWQKKTLKWPSTENYV-----KGVLLGQVNSFLINEDP--KHHPNERS 175
QY 189 GAGP-----VILPDRHYLSTQSAQSKDNEKHEKVEVLEFYAALINIMDELK 240
DB 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 176 TVGAKKGVLPVPEHYV-----GRIK-----LSHKKYNTVEVVE 311

```

RESULT 15

```

Q876C0 PRELIMINARY: PRT: 236 AA.
AC Q876C0
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DI 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
OS Dendronephthya sp. SSAL-2002.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
OC Nephtheidae; Dendronephthya.
GX NCBI_TaxID=191210;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEELINE-21927629; PubMed-11929995;
RA Lebas Y.A., Gaiskaya N.G., Yanshevich Y.G., Fradkov A.F.,
RA Lukyanov K.A., Lukyanov S.A., Malt M.V.,
RT "Diversity and evolution of the green fluorescent protein family."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DB EMBL: AF420591; AAM10025.1;
SQ SEQUENCE 236 AA: 26840 MW: CGL7075FF934A90 CFC64.

Query Match 19.6%; Score 249.5; DB 5; Length 236;
Best Local Similarity 31.8%; Pred. No. 3.5e-17;
Matches 64; Conservative 39; Mismatches 25; Indels 13; Gaps 7;

QY 14 TAVRIDGDVNGHKESVROGEGQADYVQKLEIRFICITISKLVIMPTL 60
DB 10 VKVMGQVNGHAFVIEGEGSKRPVETLTNLTVKGAPIPHSYLLITLALHYGNPVFI 65
QY 75 KYEDHMKRHDFKSAPEYVQERTIFFKQDQNYKTRASVREJILY 120
DB 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 70 EYFADIT--DYFKQSPPEYNSWERSKREAGVTVTSIKLEIDGTFYETKRYGVN 127
QY 133 DGNILGKHKLEYNNNSHNVYIMAKKQKSKVNFKIDNLEQNVLAG 190
DB 126 NGPVMQKKTLKWPSTENYV-----KGVLLGQVNSFLINEDP--KHHPNERS 175
QY 191 GPVLPDRHYLSTQSAQSKDNEKHEKVEVLEFYAALINIMDELK 240
DB 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 181 KVGLPDPVPEHYV-----GRIK-----LSHKKYNTVEVVE 311

```

Search completed: July 16, 2003, 16:13:29
 Job time : 41.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:13:38 : Search time 29 seconds
(without alignments)
974,649 Million CPU updates/sec

Title: US-09-967-301-2

Perfect score: 1275

Sequence: 1 MSKGELFTGVVPIVELDGL.....VLIEFVIAAGI-HGMDELYK 238

Scoring table: RLOSUM62

Gapop 10.0 / Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying the chosen parameters: 11899

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 1%

Maximum Match 10%

Listing: first 45 summaries

Database : published_applications.dat

1: /cgn2_6/prodata/2/pubpa/US07_NEW_PUB.pep.*
2: /cgn2_6/prodata/2/pubpa/PUB_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpa/US06_PUB_NEW_PUB.pep.*
5: /cgn2_6/prodata/2/pubpa/ACTIS_PUBCOMB.pep.*
6: /cgn2_6/prodata/2/pubpa/US07_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpa/US09_NEW_PUB.pep.*
10: /cgn2_6/prodata/2/pubpa/US09_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpa/US10_NEW_PUB.pep.*
12: /cgn2_6/prodata/2/pubpa/US10_PUBCOMB.pep.*
13: /cgn2_6/prodata/2/pubpa/US6_NEW_PUB.pep.*
14: /cgn2_6/prodata/2/pubpa/US6_PUBCOMB.pep.*

pred. No. is the number of results predicted by engine to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	1275	100.0	238	9	US-10-057-505-2
2	1275	100.0	238	10	US-09-884-561-2
3	1275	100.0	238	12	US-10-024-665-2
4	1271	99.7	238	9	US-09-866-535-2
5	1271	99.7	238	9	US-09-900-345A-120
6	1271	99.7	238	9	US-10-121-358-10
7	1271	99.7	238	9	US-10-221-451-6
8	1271	99.7	238	10	US-09-920-522-4
9	1271	99.7	238	10	US-09-852-035-1
10	1268	99.5	243	9	US-09-900-345A-60
11	1268	99.5	243	9	US-09-900-345A-62
12	1268	99.5	243	9	US-09-900-345A-64
13	1268	99.5	243	9	US-09-900-345A-66
14	1268	99.5	243	9	US-09-900-345A-68
15	1268	99.5	243	9	US-09-900-345A-70
16	1268	99.5	1070	9	US-10-001-446B-2
17	1268	99.5	1099	9	US-10-259-864-4
18	1268	99.5	1147	9	US-10-259-864-5
19	1268	99.5	1452	12	US-10-050-673-2

20	1267	99.4	243	9	US-09-900-345A-54
21	1267	99.4	243	9	US-09-900-345A-56
22	1267	99.4	243	9	US-09-900-345A-58
23	1267	99.4	243	9	US-09-900-345A-114
24	1267	99.4	243	9	US-09-900-345A-116
25	1267	99.4	243	9	US-09-900-345A-118
26	1267	99.4	243	9	US-09-900-345A-120
27	1266	99.3	243	9	US-09-900-345A-2
28	1266	99.3	243	9	US-09-900-345A-4
29	1266	99.3	243	9	US-09-900-345A-6
30	1266	99.3	243	9	US-09-900-345A-8
31	1266	99.3	243	9	US-09-900-345A-10
32	1266	99.3	243	9	US-09-900-345A-12
33	1266	99.3	243	9	US-09-900-345A-14
34	1266	99.3	243	9	US-09-900-345A-16
35	1266	99.3	243	9	US-09-900-345A-18
36	1266	99.3	243	9	US-09-900-345A-20
37	1266	99.3	243	9	US-09-900-345A-22
38	1266	99.3	243	9	US-09-900-345A-24
39	1266	99.3	243	9	US-09-900-345A-26
40	1266	99.3	243	9	US-09-900-345A-28
41	1266	99.3	243	9	US-09-900-345A-30
42	1266	99.3	243	9	US-09-900-345A-32
43	1266	99.3	243	9	US-09-900-345A-34
44	1266	99.3	243	9	US-09-900-345A-36
45	1266	99.3	243	9	US-09-900-345A-38

ALIGNMENTS

RESULT 1
US-10-057-505-2
? Sequence 2: Application US/10057505
? Patent No. US602018467JAI
? GENERAL INFORMATION:
? APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
? APPLICANT: AURORA BIOSCIENCES CORPORATION
? APPLICANT: TSIEH, Roger
? APPLICANT: HEIM, Roger
? APPLICANT: CURTIS, Andrew
? TITLE OF INVENTION: TANDEM FLUORESCENT PROTEIN CONSTRUCTS
? FILE REFERENCE: PEGN1260-3
? CURRENT APPLICATION NUMBER: US/10/057.505
? PRIORITY FILING DATE: 2002-01-25
? PRIOR APPLICATION NUMBER: US 08/792.553
? PRIORITY FILING DATE: 1997-01-31
? PRIOR APPLICATION NUMBER: US 09/596.003
? PRIORITY FILING DATE: 1999-09-13
? NUMBER OF SEQ ID NOS: 31
? SOFTWARE: Patentin version 3.1
? SEQ ID NO: 2
? LENGTH: 239
? TYPE: PPT
? ORGANISM: Aequorea victoria
US-10-057-505-2

Query Match	100.0%	Score 1275	DP 9	Length 238
Best Local Similarity	100.0%	Prod. No. 1.50-114		
Matches 238	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MSKGELFTGVVPIVELDGLGVNGHKFSVSGEGSDATYGKLTGKTFCTGKLFVPMPTL	60	
DB	1	MSKGELFTGVVPIVELDGLGVNGHKFSVSGEGSDATYGKLTGKTFCTGKLFVPMPTL	60	
QY	61	VTFYGVQCFSSRYPUHKMRHDFEKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV	120	
DB	61	VTFYGVQCFSSRYPUHKMRHDFEKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV	120	
QY	121	NRIELKGIDFKEDGKILGHKLEYNTNSNYIMADKQNGKIVNFKIRHNEIDGSGVQLAD	180	
DB	121	NRIELKGIDFKEDGKILGHKLEYNTNSNYIMADKQNGKIVNFKIRHNEIDGSGVQLAD	180	

QY 181 HYQNTPTGCGVLLDNDYVLTQSALSKDPNEKREHMYVLEFFVTAAGITHGMDELYK 238
 |||||.....|||||...|||||.....|||||.....
 Db 181 HYQNTPTGCGVLLDNDYVLTQSALSKDPNEKREHMYVLEFFVTAAGITHGMDELYK 238

RESULT 2

US-09-884-681-2

: Sequence 2, Application US/9984681

: Patent No. US2002061546A1

: GENERAL INFORMATION:

: APPLICANT: Islen, Roger Y.

: ATTORNEY/AGENT INFORMATION:

: NAME: Cobble, Andrew B.

: TITLE OF INVENTION: Assays for Protein Kinases Using

: Fluorescent Protein Substrates

: NUMBER OF SEQUENCES: 48

: CORRESPONDENCE ADDRESS:

: ADDRESSES: Townsend and Townsend and Clark LLP

: STREET: Two Embarcadero Center, Eighth Floor

: CITY: San Francisco

: STATE: California

: COUNTRY: USA

: ZIP: 94111-3814

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/884,681

: FILING DATE: 19-Jun-2001

: CLASSIFICATION: <Unknown>

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 08/673,865

: FILING DATE: <Unknown>

: ATTORNEY/AGENT INFORMATION:

: NAME: Storella, John S.

: REGISTRATION NUMBER: 32,944

: REFERENCE/DOCKET NUMBER: 023072-059000

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (415) 576-0200

: TELEFAX: (415) 576-0300

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 238 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

: SEQUENCE DESCRIPTION: SFO ID NO: 2

US-09-884-681-2

Query Match: 100.0%; Score 1275; DB 1; Length 238;

Best Local Similarity 100.0%; Pred. No. 1,5e-114;

Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELTGVVPLVLELDGVNGHKSFSVSGEGSDATYKLTIKTCTTCKLVPWPTL 60

|||.....|||||.....|||||.....|||||.....

Db 1 MSKGEELTGVVPLVLELDGVNGHKSFSVSGEGSDATYKLTIKTCTTCKLVPWPTL 60

|||.....|||||.....|||||.....|||||.....

QY 61 VTTFSGVQCFSRYPDHMKRHDFEKSAMTSGYVQERTIFPKLGNKTRAEVKFEEDTLV 120

|||.....|||||.....|||||.....|||||.....

Db 61 VTTFSGVQCFSRYPDHMKRHDFEKSAMTSGYVQERTIFPKLGNKTRAEVKFEEDTLV 120

QY 121 NRLEKLGIDFKEDGNILGHKLEYNYNHNYIMADKKNKIKYKFKIRHNIEIGSVQAD 180

|||.....|||||.....|||||.....|||||.....

Db 121 NRLEKLGIDFKEDGNILGHKLEYNYNHNYIMADKKNKIKYKFKIRHNIEIGSVQAD 180

QY 181 HYQNTPTGCGVLLDNDYVLTQSALSKDPNEKREHMYVLEFFVTAAGITHGMDELYK 238

|||.....|||||.....|||||.....|||||.....

Db 181 HYQNTPTGCGVLLDNDYVLTQSALSKDPNEKREHMYVLEFFVTAAGITHGMDELYK 238

: RESULT 3

US-10-024-666-2

: Sequence 2, Application US/0986536

: Publication No. US2003032086A1

RESULT 4

US-09-866-038-2

: Sequence 2, Application US/0986536

: Publication No. US2003032086A1

: Sequence 2, Application US/1002486
 : Patent No. US20020123113A1
 : GENERAL INFORMATION:

: APPLICANT: Islen, Roger Y.

: ATTORNEY/AGENT INFORMATION:

: NAME: Heim, Roger

: TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS

: NUMBER OF SEQUENCES: 5

: CORRESPONDENCE ADDRESS:

: ADDRESSES: Fish & Richardson P.C.

: STREET: 4225 Executive Square, Suite 1400

: CITY: La Jolla

: STATE: CA

: COUNTRY: USA

: ZIP: 92037

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: Windows95

: SOFTWARE: FastSeq for Windows Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/10/024,686

: FILING DATE: 17-Dec-2001

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 09/057,995

: FILING DATE: <Unknown>

: APPLICATION NUMBER: 08/727,452

: FILING DATE: 10-OCT-1996

: APPLICATION NUMBER: US95/14692

: FILING DATE: 13-NOV-1995

: APPLICATION NUMBER: 08/337,915

: FILING DATE: 10-NOV-1994

: ATTORNEY/AGENT INFORMATION:

: NAME: Hallie, Lisa A.

: REGISTRATION NUMBER: 38,347

: REFERENCE/DOCKET NUMBER: 07257/032002

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 619/678-5070

: TELEFAX: 619/678-5099

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 238 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

: FRAGMENT TYPE: internal

: SEQUENCE DESCRIPTION: SFO ID NO: 2

US-10-024-666-2

Query Match: 100.0%; Score 1275; DB 12; Length 238;

Best Local Similarity 100.0%; Pred. No. 1,5e-114;

Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELTGVVPLVLELDGVNGHKSFSVSGEGSDATYKLTIKTCTTCKLVPWPTL 60

|||.....|||||.....|||||.....|||||.....

Db 1 MSKGEELTGVVPLVLELDGVNGHKSFSVSGEGSDATYKLTIKTCTTCKLVPWPTL 60

|||.....|||||.....|||||.....|||||.....

QY 61 VTTFSGVQCFSRYPDHMKRHDFEKSAMTSGYVQERTIFPKLGNKTRAEVKFEEDTLV 120

|||.....|||||.....|||||.....|||||.....

Db 61 VTTFSGVQCFSRYPDHMKRHDFEKSAMTSGYVQERTIFPKLGNKTRAEVKFEEDTLV 120

QY 121 NRLEKLGIDFKEDGNILGHKLEYNYNHNYIMADKKNKIKYKFKIRHNIEIGSVQAD 180

|||.....|||||.....|||||.....|||||.....

Db 121 NRLEKLGIDFKEDGNILGHKLEYNYNHNYIMADKKNKIKYKFKIRHNIEIGSVQAD 180

QY 181 HYQNTPTGCGVLLDNDYVLTQSALSKDPNEKREHMYVLEFFVTAAGITHGMDELYK 238

|||.....|||||.....|||||.....|||||.....

Db 181 HYQNTPTGCGVLLDNDYVLTQSALSKDPNEKREHMYVLEFFVTAAGITHGMDELYK 238

GENERAL INFORMATION:

APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
 APPLICANT: TSIEN, Roger
 APPLICANT: Campbell, Robert
 TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
 FILE REFERENCE: REGEN1530-2
 CURRENT APPLICATION NUMBER: US/09/866,538
 CURRENT FILING DATE: 2001-05-24
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 2
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Aequorea victoria
 US-09-866-538-2

Query Match 99.7% Score 1271: DB 9: Length 238:
 Best Local Similarity 99.6%: Pred. No. 3, 5e-114:
 Matches 237: Conservative 1: Mismatches 0: Indels 0: Gaps 0:
 CQ 1 MSKGEELFTGVPIVLVELDGVNKGKTSVSGEGEDATYKGLIKFKICTTCKLVPWPTL 60
 DB 1 MSKGEELFTGVPIVLVELDGVNKGKTSVSGEGEDATYKGLIKFKICTTCKLVPWPTL 60
 CQ 61 VTTFSGVQCFSRYPDHMKRHDFKKSAMPEYVQERTIFFKDDGNKYNKTRAEVKEEDTLV 120
 DB 61 VTTFSGVQCFSRYPDHMKRHDFKKSAMPEYVQERTIFFKDDGNKYNKTRAEVKEEDTLV 120
 CQ 121 NRTELKGDKEKDNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNTEGSGVOLAD 180
 DB 121 NRTELKGDKEKDNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNTEGSGVOLAD 180
 CQ 181 HYQONTPIGDSFVILPDNHVLTQSALSCKDPNEKPDHNVLEFVIAGTTHGMDLEYK 238
 DB 181 HYQONTPIGDSFVILPDNHVLTQSALSCKDPNEKPDHNVLEFVIAGTTHGMDLEYK 238
 CQ 181 HYQONTPIGDSFVILPDNHVLTQSALSCKDPNEKPDHNVLEFVIAGTTHGMDLEYK 238
 DB 181 HYQONTPIGDSFVILPDNHVLTQSALSCKDPNEKPDHNVLEFVIAGTTHGMDLEYK 238

RESULT 5

US-09-900-345A-125
 Sequence 125: Application US/09/900-345A
 Publication No. US2003003199A1
 GENERAL INFORMATION:
 APPLICANT: Frazer, Ian Hector
 APPLICANT: Zhou, Jian
 TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
 TITLE OF INVENTION: EFFICIENCY OF A CODON
 FILE REFERENCE: 10338-503
 CURRENT APPLICATION NUMBER: US/09/900-345A
 CURRENT FILING DATE: 2001-07-06
 PRIOR APPLICATION NUMBER: AU PPR078
 PRIOR FILING DATE: 1999-01-09
 PRIOR APPLICATION NUMBER: PCT/AU00/006
 PRIOR FILING DATE: 2000-01-07
 NUMBER OF SEQ ID NOS: 185
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 125
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: GFP humanized
 OTHER INFORMATION: control
 US-09-900-345A-125

Query Match 99.7% Score 1271: DB 9: Length 238:
 Best Local Similarity 99.6%: Pred. No. 3, 5e-114:
 Matches 237: Conservative 1: Mismatches 0: Indels 0: Gaps 0:
 CQ 1 MSKGEELFTGVPIVLVELDGVNKGKTSVSGEGEDATYKGLIKFKICTTCKLVPWPTL 60
 DB 1 MSKGEELFTGVPIVLVELDGVNKGKTSVSGEGEDATYKGLIKFKICTTCKLVPWPTL 60
 CQ 61 VTTFSGVQCFSRYPDHMKRHDFKKSAMPEYVQERTIFFKDDGNKYNKTRAEVKEEDTLV 120

DB 61 VTTFSGVQCFSRYPDHMKRHDFKKSAMPEYVQERTIFFKDDGNKYNKTRAEVKEEDTLV 120
 CQ 121 NRTELKGDKEKDNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNTEGSGVOLAD 180
 DB 121 NRTELKGDKEKDNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNTEGSGVOLAD 180
 CQ 181 HYQONTPIGDSFVILPDNHVLTQSALSCKDPNEKPDHNVLEFVIAGTTHGMDLEYK 238
 DB 181 HYQONTPIGDSFVILPDNHVLTQSALSCKDPNEKPDHNVLEFVIAGTTHGMDLEYK 238

RESULT 6

US-10-121-258-10
 Sequence 10: Application US/0121258
 Publication No. US20030059835A1
 GENERAL INFORMATION:
 APPLICANT: Tsién, Roger
 APPLICANT: Campbell, Robert
 TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
 TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
 FILE REFERENCE: US083.102401
 CURRENT APPLICATION NUMBER: US/10/21,258
 CURRENT FILING DATE: 2002-04-10
 PRIOR APPLICATION NUMBER: 09/794,308
 PRIOR FILING DATE: 2001-02-26
 PRIOR APPLICATION NUMBER: 09/866,538
 PRIOR FILING DATE: 2001-05-24
 NUMBER OF SEQ ID NOS: 78
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 10
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Aequorea victoria
 US-10-121-258-10

Query Match 99.7% Score 1271: DB 9: Length 238:
 Best Local Similarity 99.6%: Pred. No. 3, 5e-114:
 Matches 237: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

CQ 1 MSKGEELFTGVPIVLVELDGVNKGKTSVSGEGEDATYKGLIKFKICTTCKLVPWPTL 60
 DB 1 MSKGEELFTGVPIVLVELDGVNKGKTSVSGEGEDATYKGLIKFKICTTCKLVPWPTL 60
 CQ 61 VTTFSGVQCFSRYPDHMKRHDFKKSAMPEYVQERTIFFKDDGNKYNKTRAEVKEEDTLV 120
 DB 61 VTTFSGVQCFSRYPDHMKRHDFKKSAMPEYVQERTIFFKDDGNKYNKTRAEVKEEDTLV 120
 CQ 121 NRTELKGDKEKDNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNTEGSGVOLAD 180
 DB 121 NRTELKGDKEKDNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNTEGSGVOLAD 180
 CQ 181 HYQONTPIGDSFVILPDNHVLTQSALSCKDPNEKPDHNVLEFVIAGTTHGMDLEYK 238
 DB 181 HYQONTPIGDSFVILPDNHVLTQSALSCKDPNEKPDHNVLEFVIAGTTHGMDLEYK 238

RESULT 7

US-10-221-451-6
 Sequence 6: Application US/1022-461
 Publication No. US20030092902A1
 GENERAL INFORMATION:
 APPLICANT: Marsh, Donald J.
 TITLE OF INVENTION: MELANIN CONCENTRATING HORMONE RECEPTOR
 TITLE OF INVENTION: CHIMERIC AND FUSION PROTEINS
 FILE REFERENCE: 20652p
 CURRENT APPLICATION NUMBER: US/10/221,461
 CURRENT FILING DATE: 2002-09-12
 PRIOR APPLICATION NUMBER: PCT/US01/08071
 PRIOR FILING DATE: 2001-03-14
 PRIOR APPLICATION NUMBER: 60/189,698
 PRIOR FILING DATE: 2000-03-15
 NUMBER OF SEQ ID NOS: 37


```

Query Match      99.5%; Score 1268; DB 9; Length 243;
Best Local Similarity 99.2%; Pred. No. 7c-114;
Matches 236; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVELDGDVNGHKFSVSGSGEGDATTYKLTIKFICTTGKLPVWPPTL 60
DB 6 LSKGEELFTGVVPIVELDGDVNGHKFSVSGSGEGDATTYKLTIKFICTTGKLPVWPPTL 65
QY 61 VTFSYGVOCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTFSYGVOCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 125
QY 121 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKIRHNLEDSGVQLAD 180
DB 126 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKIRHNLEDSGVQLAD 185
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYK 248
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYK 243

RESULT 11
US-09-900-345A-62
; Sequence 62, Application: US/09/900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10336-5CS
; CURRENT APPLICATION NUMBER: US/09/900345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP80678
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Loc(CTG)50FF
US-09-900-345A-62

Query Match      99.5%; Score 1268; DB 9; Length 243;
Best Local Similarity 99.2%; Pred. No. 7c-114;
Matches 236; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVELDGDVNGHKFSVSGSGEGDATTYKLTIKFICTTGKLPVWPPTL 60
DB 6 LSKGEELFTGVVPIVELDGDVNGHKFSVSGSGEGDATTYKLTIKFICTTGKLPVWPPTL 65
QY 61 VTFSYGVOCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTFSYGVOCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 125
QY 121 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKIRHNLEDSGVQLAD 180
DB 126 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKIRHNLEDSGVQLAD 185
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYK 248
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYK 243

RESULT 12
US-09-900-345A-64
; Sequence 64, Application: US/09/900345A
; Publication No. US20030031999A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-5US
; CURRENT APPLICATION NUMBER: US/99/900345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Loc(CTG)50FF
US-09-900-345A-64

Query Match      99.5%; Score 1269; DB 9; Length 243;
Best Local Similarity 99.2%; Pred. No. 7c-114;
Matches 236; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVELDGDVNGHKFSVSGSGEGDATTYKLTIKFICTTGKLPVWPPTL 60
DB 6 LSKGEELFTGVVPIVELDGDVNGHKFSVSGSGEGDATTYKLTIKFICTTGKLPVWPPTL 65
QY 61 VTFSYGVOCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTFSYGVOCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 125
QY 121 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKIRHNLEDSGVQLAD 180
DB 126 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKIRHNLEDSGVQLAD 185
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYK 248
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYK 243

RESULT 13
US-09-900-345A-66
; Sequence 66, Application: US/09/900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-5CS
; CURRENT APPLICATION NUMBER: US/09/900345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Loc(CTI)50FF
US-09-900-345A-66

Query Match      99.5%; Score 1268; DB 9; Length 243;
Best Local Similarity 99.2%; Pred. No. 7c-114;
Matches 236; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```


GenCore version 5.1.4
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: July 16, 2003, 16:07:32 : Search time 20.5667 seconds
(without alignments)
1107.035 Million cell updates/sec

Title: US-09-967-301-2
Perfect score: 1275
Sequence: 1 MSKGEELEFGVVPILVLELDS.....VILGFVTAAIITRMDEIVK 218

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 263224 seqs, 6018422 residues 263224
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_731*
1: pir1*
2: pir2*
3: pir3*
4: pir4*

Fred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the best hit found.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
1	1265	99.2	238 1	green-fluorescent
2	1201	7.2	785 2	hypothetical prote
3	9515	7.5	1241 2	coatomer complex a
4	95	7.5	336 2	hypothetical prote
5	91.5	7.2	887 2	leucyl-tRNA synth
6	88.5	6.9	861 2	leucyl-tRNA liase
7	87.5	6.9	862 2	leucyl-tRNA synth
8	87.5	6.9	2573 2	hypothetical prote
9	87	6.9	281 2	hypothetical prote
10	87	6.9	357 2	hypothetical prote
11	87	6.8	471 2	RNA (uracil-5-)-m
12	87	6.8	632 2	hypothetical prote
13	86	6.7	2314 2	cell wall-associat
14	85.5	6.7	836 1	DNA-directed DNA p
15	85.5	6.7	889 2	inter-alpha-trypsi
16	85	6.7	347 2	hypothetical prote
17	84.5	6.6	370 2	iron-sulfur cofact
18	84.5	6.6	425 2	hypothetical prote
19	84.5	6.6	564 2	AsC-type transport
20	84.5	6.6	797 2	protective surface
21	84.5	6.6	808 2	protective surface
22	84	6.6	529 2	hypothetical prote
23	84	6.6	760 2	protein F1086.14 i
24	83.5	6.5	461 2	photosystem II chl
25	83.5	6.5	616 2	hypothetical prote
26	83.5	6.5	1134 2	tRNA P ₂ receptor pr
27	83.5	6.5	1164 1	tRNA P ₂ receptor pr
28	83.5	6.5	2222 1	DNA-directed RNA p
29	83	6.5	207 2	hypothetical prote

ALIGNMENTS

Result: 1
CO1314

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C>Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text-change 23-Mar-2001
C:Accession: J50692; J01534; PQ0335; S48693; S51330; S51331
K:Prasher, D.C.; Packeridge, V.K.; Ward, W.W.; Prasher-Gast, F.G.; Cormier, M.J.
Gene 111: 229-233, 1992

A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: J01534; MUID:92175527; PMID:1347277

A:Accession: J50692

A:Molecule type: DNA

A:Residues: 1-107; S: 109-238 <PRA1>

A:Cross-references: GB:M62654; NID:q155662; PIDN:AAA27722.1; PID:q155663

A:Accession: CO1514

A:Molecule type: mRNA

A:Residues: 1-99; F: 101-140; S: 142-218; V: 220-238 <PRA2>

A:Cross-references: GB:M62653; NID:q155660; PIDN:AAA27722.1; PID:q155661

A:Accession: PQ0335

A:Molecule type: protein

A:Residues: 46-54; 74-122; 132-151; 154-183; 185-200 <PRA3>

A:Accession: S48693

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24; G: 26-156; P: 159-171; K: 173-238 <INP>

A:Cross-references: GB:L29345; NID:q606183; PIDN:AAA58246.1; PID:q606184

K:Watkins, J.N.; Campbell, A.K.

Submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Molecule type: mRNA

A:Residues: 1-13; V: 15-24; G: 26-44; N: 46-153; G: 155-156; P: 158-171; K: 173-227;

A:Cross-references: EMBL:X83909; NID:q634008; PIDN:CAA56789.1; PID:q634009

A:Experimental source: clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24; O: 26-29; R: 31-63; C: 65-103; G: 105-156; P: 158-171; K: 173-208;

A:Cross-references: EMBL:X83960; NID:q634010; PIDN:CAA58790.1; PID:q634011

A:Experimental source: clone gfp2

K:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Notes: annotation: X-ray crystallography, 1.9 angstroms, residues A' 2-79, R' 8

A:Note: engineered sequence based on J01514, cloned and expressed in Escherichia coli

K:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

A:Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

synaptosamin opes
fructose-bisphosph
hypothetical prote
hypothetical prote
vegetative incompe
nitrogenase (EC 1.
oligonucleotideph
DNA topoisomerase
probable iron-cont
hypothetical prote
dihydrolipoamide d
alpha-amylase homo
proteoglycan link
conserved hypothet
calpain (EC 3.4.22
DNA-directed DNA p

A:Contents: annotation: X-ray crystallography, 2.5 angstroms
 C:Comment: This protein is excited by the photolabile amino acid (see F16:AD:FNK) emitting
 C:Comment: The chromophore of this protein is formed by modification of Ser-244-Pro-Tyr-
 C:Genetics:

A:Gene: GFP
 A:Introns: 69/3: 167/3
 C:Superfamily: green-fluorescent proteins
 C:Keywords: chromoprotein; luminescence
 F:55-67/Cross-link: 5-maleimidylsuccinimide (Ser-Gly) #status experimental
 F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match: 99.2% Score 1265; DB 1; Length 238;
 Best Local Similarity 98.3%; Pred. No. 1; Se-24;
 Matches 234; Conservative 4; Mismatches 0; Indels 0; Gaps 0

QY 1 MSKGEELFGVVPVILVELDGVNGKPSVSSGHEENATYKTIKFKITCKIMPVNITL 6
 DB 1 MSKGEELFGVVPVILVELDGVNGKPSVSSGHEENATYKTIKFKITCKIMPVNITL 6
 QY 61 VTIESGVGVCSRYPTNKKRBUHFFKSMPEVGVGKTIFFKDSYKTAARVKGSGITLV 123
 DB 61 VTIESGVGVCSRYPTNKKRBUHFFKSMPEVGVGKTIFFKDSYKTAARVKGSGITLV 123
 QY 121 NRIELKGLDFKEDGNILGHKLEYNNSHNVYMAQKQKNGIKYKFKIRNKEDGSVLAD 183
 DB 121 NRIELKGLDFKEDGNILGHKLEYNNSHNVYMAQKQKNGIKYKFKIRNKEDGSVLAD 183
 QY 181 HYQONTPIGCGVLLPONHYLSTQSALSQDKRKHQVLLGVVTAANTTHGMDSYK 236
 DB 181 HYQONTPIGCGVLLPONHYLSTQSALSQDKRKHQVLLGVVTAANTTHGMDSYK 236

RESULT 2
 H72228
 hypothetical protein TM1624 - Thermotoga maritima (strain MS8)
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: H72228
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.; Dodson, R.J.; Haft, D.H.; Hickox
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, G.A.; Richardson, D.
 C.M.
 Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72200; PMID:9427456; PMID:11360671
 A:Accession: H72228
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1785 <ASN>
 A:Cross-references: GR:AE001896; GR:AE000512; NID:94962196; PIR:AA046691; PIR:G498213
 A:Experimental: source: strain MS8
 C:Genetics:

A:Gene: TM1624
 Query Match: 7.9% Score 101; DB 2; Length 245;
 Best Local Similarity 20.9%; Pred. No. 1.9;
 Matches 40; Conservative 27; Mismatches 64; Indels 6; Gaps 5

QY 2 SKGEELFGVVPVILVELDGVNGKPSVSSGHEENATYKTIKFKITCKIMPVNITL 61
 DB 15 NGRFSECTVPGVQVAD-----NEDLFKEIDREWIYERFESKQWKESGVLFVPSVSTLS 85
 QY 62 TIESGVGVCSRYPTNKKRBUHFFKSMPEVGVGKTIFFKDSYKTAARVKGSGITLV 123
 DB 47 -----NEDLFKEIDREWIYERFESKQWKESGVLFVPSVSTLS 85
 QY 122 RIELKGLDFKEDGNILGHKLEYNNSHNVYMAQKQKNGIKYKFKIRNKEDGSVLAD 183
 DB 89 DVTLENGVYV---GSTEDMFIEYFDVNTLV-----KKNILKVKIK-----STDPVKK 134
 QY 182 YQNTPIGCGP 192
 DB 135 LEQNYGVIGGP 145

RESULT 3

EFHC4H
 coarctate complex alpha chain homolog - human
 N:Alternate names: HEP-COP; xenopsin homolog; xenopsin-related peptide precursor
 N:Contains: xenin 25; xenopsin-related peptide
 C:Species: Homo sapiens (man)
 C:Date: 10-May-1996 #sequence_revision 06-Nov-1996 #text_change 21-Jul-2000
 C:Accession: J04668; A44317
 K:Chow, V.T.K.; Quek, H.H.
 Gene 169, 223-227, 1996

A:Title: HEP-COP, a novel human gene whose product is highly homologous to the alpha-

A:Reference number: J04668; KJ10:96194906; PMID:8647451
 A:Accession: J04668
 A:Molecule type: mRNA
 A:Residues: 1-1224 <CRG>
 A:Cross-references: GR:024105; NID:9138873; PIR:AA070679; PIR:Q1002369
 A:Experimental source: Hep4b hepatocellular carcinoma cell
 K:Feurle, G.E.; Hamscher, G.; Kistek, R.; Meyer, H.E.; Metzger, J.W.
 J. Biol. Chem. 267, 22305-22309, 1992

A:Title: Identification of xenin, a xenopsin-related peptide, in the human gastric mu-

A:Reference number: A44317; X010:93054515; PMID:1429581
 A:Accession: A44317

A:Molecule type: Protein
 A:Residues: 1-23 <PEP>
 A:Experimental source: gastric mucosa
 A:Note: sequence extracted from NCBI database (NCBI:117918)
 A:Note: plasma levels of xenin 25, as determined by immunoassay, rise after meals and

C:Genetics:
 A:Gene: GDB:COXA; HEP-COP
 A:Cross-references: GDB:4642787; OMIM:601924
 A:Map position: 1q23-1q25
 C:Superfamily: yeast coarctate complex alpha chain; WD repeat homology
 C:Keywords: duplication; hormone; plasma; stomach
 F:1-25/Product: xenin 25 #status experimental <XNP>
 F:15-38/Domain: WD repeat homology <WD1>
 F:17-25/Product: xenopsin-related peptide #status predicted <XRP>
 F:47-80/Domain: WD repeat homology <WD2>
 F:89-122/Domain: WD repeat homology <WD3>
 F:131-164/Domain: WD repeat homology <WD4>
 F:201-234/Domain: WD repeat homology <WD5>
 F:245-278/Domain: WD repeat homology <WD6>

Query Match: 7.5% Score 95.5; DB 1; Length 1224;
 Best Local Similarity 24.6%; Pred. No. 9.7;
 Matches 48; Conservative 30; Mismatches 76; Indels 4; Gaps 9

QY 62 TIESGVGVCSRYPTNKKRBUHFFKSMPEVGVGKTIFFKDSYKTAARVKGSGITLV 123
 DB 595 TEERFKALNKKYDEVLEKRYNNAKLKVCSTIAYLQKKGYPEVALHFKYQD---KTFEFL 652
 QY 113 KFFGDTNKRLEKGTDFKEDGNILGHKLEYNNSHNVYMAQKQKNGIKYKFKIRNK 165
 DB 652 ALQGNIEIAPAAKALDQKKNWKLSEVALLQGNHQTVMCYQTKNFKVSYLNLITG 712
 QY 165 ---KIKENIEDSVQ---LADHYQONTPIFGGVLLPONHYLSTQSALSQDKRKHQVLL 220
 DB 222 NLEKLNKKNKIAIRKQKMSGHYQNALYSG-----VSEVRILKNQSGKS----- 756
 QY 221 LEFVTAAGTTHGMD 235
 DB 757 LAYITAA--THGLGE 769

RESULT 4

C04466
 hypothetical protein MJ1348 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: C04466
 R:Salt, C.; White, C.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Saiton, G.G.; Black
 Kirsch, C.J.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Merrick, J.M.; Gloeck,
 rson, J.D.; Sadow, P.W.; Hanta, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996
A:Authors: Raife, B.P.; Borodovsky, M.; Kleuk, H.P.; Fraser, C.M.; Smith, R.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MIMD:96357999; PMID:8668187
A:Accession: C64468
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-336 kb
A:Cross-references: GR:U67574; GB:L77127; NID:g1591578; PIDN:AAH9460.1; PID:g1591590; T
C:Genetics:
A:Map position: FOR1295121-1296133
A:Start codon: TTG

Query Match 7.5% Score 95; DB 2; Length 336;
Best Local Similarity 25.3%; Pred. No. 2;
Matches 67; Conservative 35; Mismatches 31; Indels 82; Gaps 17;
QY 19 DGSVNHKFSVSGE-----GGGATYCKL--LAKFICTTCKLPVW----- 57
DB 91 DGDY-----YNLSGELSTIASIFAKIGKLBTRKFLNSGSE---NTYNDVCKANSEKIL 142
QY 58 -PLVATTESY--GVQFSESYPOHMK-----HGFKSAKPSYVGERTIEPKCG 104
DB 143 KSVCFDFESTERKEILNREP-ELRKDFENLYNFSDEFTTH-----MFTTNG 192
QY 105 NYKIRAFVKE-----GDTLVNRIELKIDFKEDNTL---GKLEYRYNSHNVI 152
DB 133 NKRRFLVEVEFKKIKSQISNEIVNRI-IRKFI-KNSILLALAHLLKENVEKTLIV 250
QY 153 MAUK-----GKNIKVNFKEHNKEGVSOLAHYQNTLIG:GSPVLP-DNHY 260
DB 251 MEFKEYFDLNEPNEFKNTINLTVDIATLNKGNKKEKLNFLKPIYKEI-KEPUNTY 310
QY 201 LSTGSAISKID-PNEKRRCHMYLLEVV 224
DB 311 ---KDAHNDMLNRIQLXYIVLEFI 332

RESULT 5
E82590
leucyl-tRNA synthetase XE2296 (imported) - Xylella fastidiosa (strain 943c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 *sequence_revision 20-Aug-2000 *text_change 02-Sep-2000
C:Accession: E82590
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-
Nature 406, 351-357, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MIMD:20459717; PMID:11540113
A:Note: for a complete list of authors see reference number A59426 below
A:Accession: E82590
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-887 <STM>
A:Cross-references: GR:AE004301; GB:AB003849; NID:g9910709; PIDN:AAH9475.1; GS:DB:GN001
A:Experimental source: strain 943c
R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Alvares, E.; Alvares, E.; A
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carriaro, D.M.; Curriel, B
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; From
J.D.; Jungueira, M.L.; Kemper, E.B.; Kitajima, E.P.; Krieger, J.E.; Kuramae, E.B.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Mariano, G.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Monck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J.; de Rosa, J.R.; V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2176
C:Superfamily: leucine-tRNA ligase

Query Match 7.2% Score 91.5; DB 2; Length 887;
Best Local Similarity 22.7%; Pred. No. 14;
Matches 45; Conservative 27; Mismatches 69; Indels 57; Gaps 10;
QY 49 TIGKLPVWPVLTITFSYGVQCFSRYPDHMRHDEFFKSAMFEGYVQERTIFFKQGNKY-- 106
DB 329 TNEQLPV-KVANEVLMHYGTGAVMAVPGHDQDEF--ANKYGLPIFQVIALKEIKNOE 365
QY 107 -----KIRAEVFEQDTLVNR-TELKGTDFEKDNTLGHKLEYNSHNVI 152
DB 386 STWEPVVRDWDYADKTR---EFE---LINSAEFGDGTQDAFEVLAERF----- 429
QY 153 MADKOKNG-IKVNFIRHNIEDGVSOLADHYQNTPI-----GGGVPLLFDN 198
DB 430 ---RGRGRRVNYSLR---DWGVSQRORYNGCPPIFYICPGGAVPVPEQLPWLPEN 482
QY 199 -HYLSTQSALSADKPNKR 215
DB 483 VAFSGTSGSPIKTDPEWRK 500

RESULT 6
H64102
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Ri KW20)
N:Alternate names: leucyl-tRNA synthetase
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 *sequence_revision 18-Aug-1995 *text_change 03-Jun-2002
C:Accession: H64102
R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.; Godek, A.; Kelley, J.M.; Weidma
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Gockhagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Giehl, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MIMD:95350630; PMID:7542800
A:Accession: H64102
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-861 <TIGR>
A:Cross-references: GB:U32774; GB:J42024; NID:g1573942; PIDN:AAC22551.1; PID:g-573943
C:Superfamily: leucine-tRNA ligase
C:Keywords: aminocyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 6.9% Score 88.5; DB 2; Length 861;
Best Local Similarity 24.3%; Pred. No. 24;
Matches 45; Conservative 25; Mismatches 66; Indels 47; Gaps 9;
QY 49 TTGSKLPVWPVLTITFSYGVQCFSRYPDHMRHDEFFKSAMFEGYVQERTIFFKQGNKI 108
DB 324 TGDKLPV-KVANEVLMHYGTGAVMAVPAHDQD--FEAQKYSLPKQVATPAIDFEID 370
QY 109 RAEVFEQDTLVNRTELKGTDFEKDNTLGHKLEYNSHNVIYMAK CKNEI---KVN 164
DB 371 TQGAQVEHCKLVNSDEFGKNP--DGAENG-----IADKLEKLGKGRQVN 414
QY 165 FKLRH-----NIDGVSQGLAHYQNTPIGUGPVLPLNNHYL-STQSAL 207
DB 415 YRLRWGVSQRORYNGCPPIMLTLENDGVVEA-----PMEDCPILPDEVYKDVAKSPI 467
QY 208 SKDPN 212
DB 468 NADPN 472

RESULT 7
AC0582
leucyl-tRNA synthetase (imported) - Salmonella enterica subsp. enterica serovar Typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 *sequence_revision 09-Nov-2001 *text_change 27-Nov-2001
C:Accession: AC0582


```

QY 141 LEYNYSIN-VYIMADKKN-----GIKVNFKIRHNLDVSVLADHY-----QONTFF 188
      || || || || || || || || || || || || || || || || || || || || ||
DB 1226 ETEYNKNDVT-KKRDTEGTVTQIAYKQIUVAVSETQSGRSSAAATCKVGNCTGSKG 1285
      || || || || || || || || || || || || || || || || || || || || ||
QY 189 GDSVILLDNVLTQSALSKCPNPKR-----DEKCVLEEVFANGTCHG 242
      || || || || || || || || || || || || || || || || || || || || ||
DB 1286 SASTNLDKGSFPAKSKWNLTAQRKRKLSVIAKESVLSAKALENTLSQSTNSGIDWG 1345

RESULT 14
J0VCLD
DNA-directed DNA polymerase (EC 2.7.7.3) - duck hepatitis virus
C:Species: duck hepatitis virus, DHBV
C:Date: 20-Sep-1984 #sequence_revision 06-Nucleotide #ext_name 11 Jan-1999
C:Accession: S12844; S36514; A30710
R:Mattes, F.; Torq, S.; Teubner, K.; Baum, H.E.
Nucleic Acids Res. 18, 5140, 1990
A:Title: Complete nucleotide sequence of a German duck hepatitis B virus.
A:Reference number: S12843; M010:51045092; P010:2235627
A:Accession: S12844
A:Status: translation not shown
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-836 <MAT>
A:Cross-references: EMBL:X12738; NID:959537
A:Experimental source: isolate DHBV 53-6
A:Note: this ORF is not annotated in GenBank entry DHBV12, release 103
R:Kunishi, A.; Panda, S.K.
J. Virol. 49, 782-792, 1984
A:Title: Nucleotide sequence of a cloned duck hepatitis B virus genome: comparison with
A:Reference number: A92997; M010:84138772; P010:669938
A:Accession: A00710
A:Molecule type: DNA
A:Residues: 390-734, A, 736-836 <MAN>
A:Cross-references: GB:K01834
A:Note: only part of the sequence reported in the GenBank entry is shown in the published
C:Superfamily: hepatitis virus DNA-directed DNA polymerase
C:Keywords: DNA biosynthesis; nucleotidyltransferase

Query Match: 6.74; Score 85.5; DB 1; Length 862
Best Local Similarity 25.28; Pred. No. 40;
Matches 34; Conservative 20; Mismatches 50; Indels 3; Gaps 0;

QY 57 WPLVLTTFVGVGCHFSYVCHMKRHHFFSAPRQVYVLEPTFFKCKYNTKRAVKKFG 114
      || || || || || || || || || || || || || || || || || || || || ||
DB 189 WPKSISYFVGVGVKPKYFNWCH-----ESTV-----KYITRD...YEA 227
      || || || || || || || || || || || || || || || || || || || || ||
QY 117 DTLVNLKELGIDFKEDGNILGKLEYNYSNHNVIIMAKKRNQIKVFKIRHNIEG 175
      || || || || || || || || || || || || || || || || || || || || ||
DB 228 GILYKRIKRLVTFK--GGPNYKQOHLVNHIVDAISSKINGPQIDRRKKNVRSN 295
      || || || || || || || || || || || || || || || || || || || || ||
QY 176 VQLADHYQONTFFGG 190
      || || || || || || || || || || || || || || || || || || || || ||
DB 286 -----KKNDPKED 293
      || || || || || || || || || || || || || || || || || || || || ||

RESULT 15
JC5576
Inter-alpha-trypsin inhibitor heavy chain A - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #ext_name 17-Mar-1999
C:Accession: JC5576; PC4486
R:Nakatani, T.; Suzuki, Y.; Yamamoto, T.; Sinohara, H.
J. Biochem. 122, 71-82, 1997

```

```

A:Title: Molecular cloning and sequencing of cDNAs encoding three heavy-chain precurs
sin inhibitor heavy chain family.
A:Reference number: JC5574; M010:97420686; P010:9276673
A:Accession: JC5576
A:Molecule type: mRNA
A:Residues: 1-869 <NAK>
A:Cross-references: DDBJ:U89287
A:Experimental source: liver
A:Accession: PC4486
A:Molecule type: protein
A:Residues: 34-53; 49-475; 509-526 <NA2>
C:Comment: In the plasma three inter-alpha trypsin inhibitor heavy chains 1, 2 and 3
that the complexes play important role for pancreatic cancer.
C:Superfamily: inter-alpha-trypsin inhibitor complex component 11
C:236-239, 664-865/Disulfide bonds: #status predicted

Query Match: 6.74; Score 85.5; DB 2; Length 869
Best Local Similarity 23.28; Pred. No. 44;
Matches 36; Conservative 34; Mismatches 62; Indels 13; Gaps 7;

QY 74 YPDHMKRHDFKSMFEGGVVQBEF-----LFFKDDGNYKIRAEVKEFGS---ILVNRIF 124
      || || || || || || || || || || || || || || || || || || || || ||
DB 479 YPEKAIL-DLTKNSYIHFDGSETAVAGRLADSDNNFK--ADVKHGALKLTLTTFEVD 535
      || || || || || || || || || || || || || || || || || || || || ||
QY 125 LKQID--FKEDGNILGKLEYNYSNHNVIIMAKKRNQIKVFKIRHNIEGSLVLAHY 182
      || || || || || || || || || || || || || || || || || || || || ||
DB 536 MKENDAAKKEQGVIFGNYIERLWAVLTIEQLLEKRN---ANGEEKENITACALELSKY 592
      || || || || || || || || || || || || || || || || || || || || ||
QY 183 QQNIP-GDSVPVLLPNNHYLSLSALSKEPNE 213
      || || || || || || || || || || || || || || || || || || || || ||
DB 593 HFVITPLPMVWIKPELN--EDQTSIAKPKGE 621
      || || || || || || || || || || || || || || || || || || || || ||

Search completed: July 16, 2003, 16:14:37
Job time : 22.6667 secs

```


GenCode Version 5.1.5
Copyright (c) 1993 - 2003 Comptech Ltd.

OM protein - protein search, using sw mode:

Run on: July 16, 2003, 10:01:23 : Search time 12.6667 seconds

(without alignments)
925.441 Million cell updates/sec

Title: US-09-067-X01-2

Perfect score: 1275

Sequence: 1 MSKGELEFGVVPILDELGG.....VLLEFVTAG:THRMCDLYK 238

Scoring table: BLOSUM62

Gapop 10.0 : Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	OM ID	Description
1	1272	99.7	238	GFP_AEQVI	P42212 Aequorea victoria (Jellyfish)
2	95.5	7.5	1224	GCPA_HUMAN	P53621 Homo sapiens
3	95	7.3	936	YDAR_MITJA	Q38743 Arabidopsis thaliana
4	92.5	7.3	1224	GCPA_BOVIN	Q27541 Bos taurus
5	91.5	7.2	879	SVL_XLIPA	Q96088 Xenopus laevis
6	88.5	6.9	863	SVL_HABIN	P43823 Haeuophilus
7	87	6.8	837	TRMA_GABCE	Q47742 Campylobacter
8	86	6.7	2734	WAPA_HACSC	Q57813 Bacillus subtilis
9	85.5	6.7	533	CP51_CANGA	Q50659 Candida glabrata
10	85.5	6.7	486	ITHI_MPSAT	P97280 Pseudomonas
11	84.5	6.6	795	U164_HAFIN	P45535 Haemaphysalis
12	84.5	6.6	797	U151_HAFIN	P35824 Haemaphysalis
13	83.5	6.5	441	PSB2_GVAPA	P43104 Cytophaga
14	83.5	6.5	733	D152_HABIN	Q32628 Haemaphysalis
15	83.5	6.5	1164	BAQ_STRAG	P27541 Streptococcus
16	83.5	6.5	2222	DIPOE_YEAST	P21351 Saccharomyces
17	83	6.5	439	SY62_DFSOM	P24506 Discomyces
18	82.5	6.5	533	N1P0_GLOPA	P60467 Cristidium
19	82.5	6.5	513	PEPF_MYCPH	Q48906 Mycoplasma
20	82	6.4	774	AMV2_SCHPO	Q43318 Schistosoma
21	81.5	6.4	355	PLK_CHECK	P07354 Gallus gallus
22	81.5	6.4	333	MTG3_HCLPV	Q25443 Malvestrum
23	81.5	6.4	700	CAN2_HUMAN	P17658 Homo sapiens
24	81.5	6.4	788	SP03_HPBRE	P13546 Heron hepatic
25	81	6.4	636	DF01_HPB09	P01162 Duck hepatitis
26	81	6.4	953	YNM2_YEAST	P53917 Saccharomyces
27	80.5	6.3	324	CAN2_PIG	P43352 Sus scrofa
28	80.5	6.3	752	NEC1_PAT	P08640 Rattus norvegicus
29	80.5	6.3	860	SVL_ECOLI	Q09313 Escherichia
30	80	6.3	297	CAN2_BOVIN	Q29571 Bos taurus
31	79.5	6.2	217	EXPI_ERWCA	P33882 Erwinia car
32	79.5	6.2	422	CAN2_RABIT	P06814 Cricetulus
33	79.5	6.2	468	GLNA_AZOCA	P94126 Azorhizobium

34	79.5	6.2	887	1	TH3_RAT	Q63416 Rattus norvegicus
35	79	6.2	504	1	YCO3_KLEPN	Q48449 Klebsiella
36	79	6.2	674	1	SIAP_BACLI	P49052 Bacillus
37	78.5	6.2	269	1	PERK_MOUSE	P22005 Mus musculus
38	78.5	6.2	538	1	GRRE_RAF	Q88500 Rattus norvegicus
39	78.5	6.2	656	1	ADAS_HUMAN	Q00116 Homo sapiens
40	78	6.1	1259	1	VTPN_ECOLI	P19321 Escherichia
41	78	6.1	2258	1	FA5_PIG	Q94101 Sus scrofa
42	77.5	6.1	312	1	TRXB_CHLMU	Q9PXL7 Chlamydia
43	77.5	6.1	505	1	SRK1_SPOJA	P42656 Spondylia
44	77.5	6.1	546	1	AMVB_RACCE	P36924 Bacillus
45	77.5	6.1	700	1	CAN2_MOUSE	Q08529 Mus musculus

ALIGNMENTS

RESULT 1	
GFP_AEQVI	STANDARD: PRT: 238 AA
AC	P42212: Q17104: (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	15-OCT-2001 (Rel. 40, Last annotation update)
DE	Green fluorescent protein.
GN	GFP.
OS	Aequorea victoria (Jellyfish).
CC	Eukaryota; Metazoa; Chordata; Hydrozoa; Hydrozoa; Leptomedusae;
CC	Aequoreidae; Aequorea.
OX	NCBI_TaxID=6100;
RN	[1]
RF	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RF	MEDLINE=92175527; PubMed=1347277;
RA	Prasher D.C., Eckhardt V.K., Ward W.W., Prendergast P.G.,
RA	Cormier M.J.;
RT	"Primary structure of the Aequorea victoria green-fluorescent
RT	protein."
RL	Gene 11:229-233(1992).
RN	[2]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=94185810; PubMed=8187953;
RA	Lucy S., Tsuji F.;
RT	"Aequorea green fluorescent protein. Expression of the gene and
RT	fluorescence characteristics of the recombinant protein."
RL	FEBS Lett. 341:277-280(1994).
RN	[3]
RP	CHROMOPORIN.
RA	MEDLINE=93192221; PubMed=8442132;
RA	Cody C.W., Prasher D.C., Webster W.M., Prendergast P.G., Ward W.W.;
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea
RT	green-fluorescent protein."
RL	Biochemistry 32:1212-1218(1993).
RN	[4]
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA	MEDLINE=96355655; PubMed=8703075;
RA	Ormeo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.;
RT	Remington S.J.;
RT	"Crystal structure of the Aequorea victoria green fluorescent
RT	protein."
RL	Science 273:1392-1395(1996).
RN	[5]
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA	MEDLINE=98294543; PubMed=9631087;
RA	Yang F., Moss L.G., Phillips G.N. Jr.;
RT	"The molecular structure of green fluorescent protein."
RT	"The molecular structure of green fluorescent protein."
RL	Nat. Biotechnol. 14:1246-1251(1996).
RN	[6]
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
RA	MEDLINE=98455509; PubMed=9782051;
RA	Wachter R.M., Elslinger M.A., Kallio K., Hanson G.T., Remington S.J.;
RT	"Structural basis of spectral shifts in the yellow-emission variants
RT	of green fluorescent protein."
RL	Structure 6:1267-1277(1998).

[7] X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 MEDLINE:92238303; PubMed:10220315;
 Eisiger M.A., Wachtel P.M., Hanson G.T., Kalish K., Remington S.J.;
 "Structural and spectral response of green fluorescent protein
 variants to changes in pH";
 Biochemistry 38:5296-5301(1999).
 -!- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSFER THE
 BLUE CHEMILUMINESCENCE OF THE PROTEIN APOPOKIN INTO GREEN
 FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON
 RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PROTEIN APOPOKIN.
 ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
 ABSORPTION PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
 PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.
 -!- SUBUNIT: MONOMER.
 -!- TISSUE SPECIFICITY: PHOTOCYTIS.
 -!- PTM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE WHICH IS COMPOSED
 OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
 CYCLIZATION OF THE RESIDUES SER-DEHYDROXYR-SER.
 -!- BIOCHEMISTRY: Has become a useful and ubiquitous tool for molecular
 cloning as a fluorescent protein tag. The tag is used to monitor
 functions as a fluorescent protein tag. The tag is used to monitor
 terminal fusion to a broad variety of proteins. It has been
 expressed in bacteria, yeast, slime mold, plants, drosophila,
 zebrafish, and in mammalian cells. As a non-invasive fluorescent
 marker in living cells, it allows for a wide range of applications
 where it may function as a cell lineage tracer, reporter of gene
 expression, or as a measure of protein-protein interactions.
 -!- DATABASE: NAME: Protein Spotlight.
 NOTE: Issue 11 of June 2001;
 WWW: <http://www.expasy.org/spotlight/articles/spl101.html>.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Made by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 PMID: 162654; AAA27722.1;
 PMID: 162653; AAA27721.1;
 PMID: 1293445; AAA58246.1;
 PDB: 1GFL; 11-JAN-97.
 PDB: 1EMA; 08-NOV-96.
 PDB: 1EMB; 16-JUN-97.
 PDB: 1PMC; 20-AUG-97.
 PDB: 1PMD; 20-AUG-97.
 PDB: 1PME; 20-AUG-97.
 PDB: 1PMF; 20-AUG-97.
 PDB: 1PMG; 12-MAY-99.
 PDB: 1PMH; 20-AUG-97.
 PDB: 1PMI; 20-AUG-97.
 PDB: 1PMJ; 20-AUG-97.
 PDB: 1PMK; 20-AUG-97.
 PDB: 1PMN; 20-AUG-97.
 PDB: 1PMO; 20-AUG-97.
 PDB: 1PMQ; 07-JUL-97.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98.
 PDB: 1PMU; 22-OCT-98.
 PDB: 1PMV; 22-OCT-98.
 PDB: 1PMW; 22-OCT-98.
 PDB: 1PMX; 22-OCT-98.
 PDB: 1PMY; 22-OCT-98.
 PDB: 1PMZ; 22-OCT-98.
 PDB: 1PMA; 22-OCT-98.
 PDB: 1PMB; 22-OCT-98.
 PDB: 1PMC; 22-OCT-98.
 PDB: 1PMD; 22-OCT-98.
 PDB: 1PME; 22-OCT-98.
 PDB: 1PMF; 22-OCT-98.
 PDB: 1PMG; 22-OCT-98.
 PDB: 1PMH; 22-OCT-98.
 PDB: 1PMI; 22-OCT-98.
 PDB: 1PMJ; 22-OCT-98.
 PDB: 1PMK; 22-OCT-98.
 PDB: 1PMN; 22-OCT-98.
 PDB: 1PMO; 22-OCT-98.
 PDB: 1PMQ; 22-OCT-98.
 PDB: 1PMR; 22-OCT-98

DB 311 ---KDAHNDLLNEIJOYVYVLEFI 302

RESULT 4

COPIA_BOVIN STANDARD: PRT: 1224 AA.
 AC Q27954;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coatamer alpha subunit (Alpha-coat protein) (Alpha-COP) (HFE008)
 DE (HEP-COP) [contains: Xenin (Xenopus-related peptide) (Protein)]
 GN COPIA.
 OS Bos taurus (Bovidae).
 OC Eukaryota; Metazoa; Chordata; Cnidaria; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Bovina; Bovidae;
 OC Bovidae; Bovinae; Bos.
 CX NCBI_TaxID:9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kiehl and F.L.
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THE COATMER IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS
 CC TO DILYSINE MOTIFS AND REVERSLY ASSOCIATES WITH GOLGI-TO-ER
 CC CLATHRIN-COATED VESICLES, WHICH FOR THEP MEDIANE BIOSYNTHETIC
 CC PROTEIN TRANSPORT FROM THE ER, VIA THE GOLGI UP TO THE TRANS GOLGI
 CC NETWORK. COATMER COMPLEX IS REQUIRED FOR HYDING FROM GOLGI
 CC MEMBRANES, AND IS ESSENTIAL FOR THE RETRANSFER COLOI-TO-ER
 CC TRANSPORT OF DILYSINE-TAGGED PROTEINS. IN MAMMALS, THE COATMER
 CC CAN ONLY BE RECRUITED BY MEMBRANES ASSOCIATED TO ARP-RIBOSYLATION
 CC FACTORS (ARPS), WHICH ARE SMALL GTP-BINDING PROTEINS; THE COMPLEX
 CC ALSO INFLUENCES THE GOLGI STRUCTURAL INTEGRITY, AS WELL AS THE
 CC PROCESSING, ACTIVITY, AND ENDOCYTIC RECYCLING OF LDL RECEPTORS (BY
 CC SIMILARITY).
 CC -!- FUNCTION: XENIN STIMULATES EXOCRINE PANCREATIC SECRETION. IT
 CC INHIBITS PENTAGASTRIN STIMULATED SECRETION OF ACID TO INHIBCE
 CC EXOCRINE PANCREATIC SECRETION AND TO AFFECT SMALL AND LARGE
 CC INTESTINAL MOTILITY. IN THE GUT, XENIN INTERACTS WITH THE
 CC NEUTROGENIN RECEPTOR (BY SIMILARITY).
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX THAT CONSISTS OF AT LEAST THE ALPHA,
 CC BETA, BETA', GAMMA, DELTA, EPSILON AND ZETA SUBUNITS.
 CC -!- SUBCELLULAR LOCATION: THE COATMER IS CYTOSOLIC OR POLYMERIZED
 CC ON THE CYTOSOLIC SIDE OF THE GOLGI, AS WELL AS ON THE
 CC VESICLES/BUDS ORIGINATING FROM IT (BY SIMILARITY). XENIN IS
 CC SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usurped, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/academic/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: X56766; CAA6543.2;
 DR Interim: F8001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR PRINTS: PR00320; GPROTHINHRPT.
 DR ProDom: PD000218; WD40; 5.
 DR SMART: SM00320; WD40; 6.
 DR PROSITE: PS00678; WD_REPEATS_2; 1.
 DR PROSITE: PS00682; WD_REPEATS_2; 6.
 DR PROSITE: PS00294; WD_REPEATS_REGION; 1.
 KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum;
 KW Membrane; Repeat; WD repeat; Hormone.
 FT PEPTIDE : 25
 FT PEPTIDE : 35 PROHENTIN.
 FT REPEAT : 7 37 WD 1.
 FT REPEAT : 49 79 WD 2.
 FT REPEAT : 91 121 WD 3.
 FT REPEAT : 133 163 WD 4.
 FT REPEAT : 203 233 WD 5.

FT REPEAT 247 277 WD 6.
 SQ SEQUENCE 1224 AA: 138358 MW: C93508P2AC00683D CRC64:
 Query Match 7.3% Score 92.5; DB 1; Length 1224;
 Best Local Similarity 24.18; Pred. No. 5.2; Mismatches 76; Indels 41; Gaps 9;
 Matches 47; Conservative 31;
 QY 62 TTESYGVGVCFSPYD---HMKRH-----DFFKSAMPEGYVQEMTIFPKDGNKTKRAEV 112
 QY 595 TEFREKLALINRKYDEVJHVNNAKINGQSLAYLQKKGYPEVALHEVKEF---KTFESL 651
 QY 113 KFEQDTLVNRTS-LKGIDFKDGNLGGKLEYNYNHNYIMADKQNGKKNV----- 165
 QY 652 ALECGNPIALEAAKALDKKNCNEKLEVALLOGNHQIVEMCYORTNFKLSELYITG 711
 QY 166 ---KIRHNEDGSVC--LADHYQNTFLGGPVLIPDNHYLSTQSALSCKINKEKDHML 720
 QY 712 NLEKPKMKKAEIFKEMSGHYQNALYAD-----VSEVRVILKNGGKS----- 756
 QY 221 IEPVTNAGITHOMDE 235
 QY 757 LAYLTAA--THGLDE 769
 RESULT 5
 SVL_XYLFA STANDARD: PRT: 873 AA.
 QY 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
 GN LEUS OR XF2176.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 CX NCBI_TaxID-2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:9a5C;
 RX MEDLINE:20365717; PubMed:10910347;
 RA Simpson A.J.G., Reirach F.C., Artuda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.B.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madris F.A.L., Martins E.M.F., Marino C.L.,
 RA Marques M.V., Martins F.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nishi A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
 RA Peixoto R.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.R.,
 RA Quadrio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchiko M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,
 RT "The genome sequence of the plant pathogen Xylella fastidiosa".
 RL Nature 406:151-155(2000).
 CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) + AMP +
 CC diphosphate + L-leucyl-tRNA(Leu).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.

[2]
 RP SEQUENCE OF 31-50: 446-472 AND 504-523, AND SUBUNIT 1.
 RC TISSUE-Plasma;
 RA MEDLINE-97018241; PubMed-8854857;
 RT Yamamoto T., Yamamoto K., Sishida H.:
 "Inter-alpha-trypsin inhibitor and its related proteins in Syrian hamster urine and plasma."
 RL J. Biochem. 120:145-152(1996).
 CC -1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SEMIN OR AS A BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN, INCLUDING THOSE ON CELL SURFACES IN TISSUES BY REGULATE THE LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY SIMILARITY).
 CC -1- SUBUNIT 1: ALPHA-1 PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN, RIKUNIN, INTER-ALPHA-INHIBITOR (I-ALPHA-I) IS COMPOSED OF H1, H2 AND BIKUNIN, INTER-ALPHA-LIKE INHIBITOR (I-ALPHA-LI) OF H2 AND BIKUNIN, AND PRE-ALPHA-INHIBITOR (P-ALPHA-I) OF H3 AND BIKUNIN.
 CC -1- PFM: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CROSSLINK
 CC 4-SULFATE BRIDGE TO THE THEIR C-TERMINAL ASPARATE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TTR FAMILY
 CC -1- SIMILARITY: CONTAINS 1 VWEA DOMAIN
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL database. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/> or send an email to license@sib-sib.ch).
 DR EMBL: D89287; AAC1340.1;
 DR InterPro: IPR002035; VWEA;
 DR Pfam: PF00092; vwa; 1;
 DR SMART: SM00327; VWA; 1;
 DR PROSITE: PS02334; VWEA; 1;
 KW Serine protease inhibitor; Repeat: Signal: Multidomain family;
 FT Glycoprotein. 1 16
 FT SIGNAL: 1 16
 FT PROPEP 19 30
 FT CHAIN 31 646
 FT PROPEP 647 866
 FT DOMAIN 279 439
 FT CARGOYD 88 88
 FT CARGOYD 577 577
 FT BINDING 545 646
 FT SEQUENCE 886 AA; 99018 MW; AC0594C6852576B CRC64;
 Query Match 6.74; Score 85.5; Db 1; Length 886;
 Best Local Similarity 23.84; Pred. No. 14;
 Matches 36; Conservative 34; Mismatches 52; Indels 19; Gaps 7;
 QY 74 YFHKRKHDFPKSAMPEYVQERT-----TEFKDGGYKRAEVKPEGD---TLVNRIE 124
 DB 476 YFENRIE-DL-KNSVPHEVYQSEFAVAKAASAMNKK-AVAKDGLNGLN-TFTPEVD 512
 QY 125 LKGIQD--FKESGNLGHKLEYNVSHRYVYDAKQKNDKYVSKTKPHN-EDYSGVLADH 144
 DB 543 KKEHMAALKEGYTFIRVYIEKMAVYITTEQLLEKPK--AADEKPKFNLAADELSHY 564
 QY 163 QNTPLDGPVLLTNNYLSQSALSKDNE 214
 DB 590 HFFVPLTPTKTKPDN--EPTSTACKPGD 616
 RESULT 12
 ID D152_HAFIN STANDARD: PRI: 795 AA.
 AC P44935;

QY 01-NOV-1995 (Rel. 32, Created)
 DB 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DB Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DB (p15-Aq) (Outer membrane protein D15).
 DB HI9917.
 DB Haemophilus influenzae.
 DB Bacteria: Proteobacteria; gamma sub-division; Pasteurellaceae;
 DB Haemophilus.
 DB NCBI_TaxID=727;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RA MEDLINE=8535033; PubMed=7542800;
 RA Fleischman R.D., Adams M.D., White O., Clayton R.A., Kirkness R.F.,
 RA Kerlavage A.R., Bitt C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McEwen K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu D.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon N.G.,
 RA Fine L.D., Fritchman J.D., Fritchman J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.J.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
 RI Science 269:466-512(1995).
 CC -1- SUPRACELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL database. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/> or send an email to license@sib-sib.ch).
 CC EMBL: U32773; AAC22575.1;
 CC TISH: H20517;
 DR InterPro: IPR000184; Bac_surface_Ag_1;
 DR Pfam: PF01103; Bac_surface_Ag; 1;
 KW Antigen; Outer membrane; Signal; Complete proteome.
 FT SIGNAL 1 19
 FT SIGNAL 20 795
 FT CHAIN 795 AA; 87478 MW; B85691P22B6FD4 CRC64;
 SC SEQUENCE 795 AA; 87478 MW; B85691P22B6FD4 CRC64;
 Query Match 6.58; Score 84.5; Db 1; Length 795;
 Best Local Similarity 21.78; Pred. No. 14;
 Matches 47; Conservative 31; Mismatches 76; Indels 63; Gaps 12;
 QY 66 YGVCFGRYPDHMKRHDFF-----PKSAMPEYVQ-----PFIIF 100
 DB 429 YGTSSGTSYQASVKODNPLGTGAAYSLASTKNXYGTSVNLGYTFYPTKUNSLGNVFP 488
 QY 101 KDDGNVTRAEVKPEGDTLVNRIELKGLDKEKGN---LGR-----KLEYNYSNR 149
 DB 489 KYNDSKSDTSNKRTYGSNVTL-GFPVNNKNSYVVGUGHTYNNKISPALEYN---RN 544
 QY 150 VYIMADKQK-NGLKVNKIRHNJEDGSGVLADHYQ-----NTYKSSVLLP- 196
 DB 545 LYTSMKFKGSHKLTN-----DEDFSGWNNINSINRGYFTTKGVKASLT-GRVING 595
 QY 597 -DNHYLSQSALSKDNEKRRHMYLLEFV-ANGITHS 232
 DB 596 SDNYYKLSADVGQFYDIDRDLHWVSAKASAGYANG 632
 RESULT 12
 ID D151_HAFIN STANDARD: PRI: 797 AA.
 AC P46024;
 DT 01-NOV-1995 (Rel. 42, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)


```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D15-Ag; Outer membrane protein D15).
CS Haemophilus influenzae.
CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
CC Haemophilus.
CX NCBI_TaxID=727;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Minnesota and Eagan / Serotype B;
RX MEDLINE=95253676; PubMed=737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RT Haemophilus influenzae."
RI Gene 156:97-99(1995).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=Minnesota and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=728140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortt D.J.K., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RI Infect Immun 65:3701-3707(1997).
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=Minnesota and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=728140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortt D.J.K., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RI Infect Immun 65:3701-3707(1997).
CC 1- SUBCELLULAR LOCATION: Outer membrane.
CC 1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announcement/
CC or send an email to license@sib-sib.ch.)
DR EMBL: U13961; AA055445.1;
DR EMBL: U60532; AB01974.1;
DR EMBL: U60533; AB01976.1;
DR InterPro: IPR00184; BacSurfAg_D15.
DR Flank: PF01103; BacSurfAg_Ag_1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL
FT CHAIN
FT 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA: 87675 MW: 259305.466 kDa: B 0904;
Query Match 6.68; Score 24.5; E: 1e-05; Identities: 70; Gaps: 11;
Best Local Similarity 20.74; Pred. No. 15;
Matches 47; Conservative 33; Mismatches 90; Indels 83; Gaps 10;
CY 66 YGVCFNRYFPHMKRHLP----- --FKSAMPPEVVE-----ELTF 17
DS 429 KATSGISQVQASVQDNFLGTGAASVATKNDAGISVAVTETFTKDVSLKVEF 486
CY 101 KKGNGKRAEVKFGGTPVNRIRKLEKEDKQNI-- --LGR-----KLEFYNSEN 147
DS 489 ENYENSKSDTSNKKATYGSNVL--GFPPNENSTYVGLDHTNKNLSNFAEYN-- --RH 544
CY 130 VYIMADKQK-- --KVKVFAIRNIEDSGVOLADHYQV----- --NTPTGQVPLD 596
DS 545 LYVQSMKFGKNGIKTN----- --DFPFGKNVNSLKNQVETKGVKASLG-- --GYTLP 545
CY 197 -DNHYLSTQSALSKDPNKKRDMKLVLLFTVANGILHG 242
DS 596 SDNKKYKLSADVQGGPYPLDQPHQWVSKAKASAGYANG 632
RESULT 13
PSBC_CVAPA STANDARD: PRT: 46; AA
AC P48104;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

```

RESULT 14

D153_HAETN STANDARD: PRT: 793 AA.
AC C32629;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (3-15-Ag) (outer membrane protein D15)
OS Haemophilus influenzae
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales
OC Haemophilus
OC NCBI_TaxID=727;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RX MEDLINE=9742792; PubMed 9284140;
RA Locmore S.M., Yang Y.F., Coleman D.C., Shortland D.K.,
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3731-3737(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U63834; AAB61977.1;
CC InterPro: IPR00184; Bac_surfAg_D15.
CC Pfam: PF01103; Bac_surface_Ag_1;
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA: 87511 MW: 51862036601A14 CRC64:

Query Match 6.5%; Score 83.5; DB 1; Length: 793;
Best Local Similarity 22.1%; Pred. No. 18;
Matches 48; Conservative 24; Mismatches 77; Indels 83; Gaps 11;
QY 66 YGVQCFSSRYPEHMKRDF-----FKSAMDESYVE-----MTTF 160
DE 429 YGTESG:SVYTS:KQNFGLTGAAVTAG:KNYH:SVNLGY:EPYFKGVSLGNIFP 498
QY 101 KDGNYKTRAEVKEFD:LNVRIELKQIDPK:KGN:---LGH-----KLEYNYSIN 149
DE 489 ENYDNRKSTSSNYKRTTYGSNVL:GFFVNNNSVYVGLGHTYKTSNFALYN:--RN 544
QY 150 VYIMAKOK:NGIKVNFKISHNEDGNSVOLADYVQ:---NTTIGGPEVLP:-- 136
DE 545 LYQSMKFKNGKTN-----DLEFSQKYNLSNRYETFKVKNASIR:DEVTTFG 545
QY 197 -DNRYLSTOSALSQRNEKTHVLEFVIAATHE 312
DE 596 SNKYYKLSADVGQGYPLDCHRWVNSAKASAVANQ 612

RESULT 15
BAG_STRAG
ID BAG_STRAG STANDARD: PRT: 1164 AA.
AC P27951;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE IgA FC receptor precursor (Beta antigen) (B antigen).
GN BAG.
OS Streptococcus agalactiae.

Bacteria; Firmicutes; Bacetobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1311;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-49.
RC STRAIN=LA239;
RX MEDLINE=913121; PubMed 1957207;
RA Jurlstrom P.G., Chhatwa G.S., Timmis K.N.,
RT "The IgA-binding beta antigen of the C protein complex of Group B
RT streptococci: sequence determination of its gene and detection of two
RT binding regions."
RL Mol. Microbiol. 9:843-849(1991).
RN (2)
RP IDENTIFICATION OF IG-LIKE DOMAIN.
RC MEDLINE=97335265; PubMed=8880921;
RA Bateman A., Pddy S.R., Chothia G.
RT "Members of the immunoglobulin superfamily in bacteria."
RX Protein Sci. 5:1939-1942(1996).
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X53771; CAA42442.1;
CC PIR: S5330; FCSAG.
CC InterPro: IPR04829; Csurface_antigen.
CC InterPro: IPR001899; Gram_pos_anchor.
CC InterPro: IPR003599; Ig.
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC ProDom: PD153432; Csurface_antigen; 2.
CC SMART: SM00409; Ig; 1.
CC TIGRFAMs: TIGR01167; LPXIG_anchor; 1.
CC TIGRFAMs: TIGR01168; YSIRK_signal; 1.
CC PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
KW Immunoglobulin domain.
FT SIGNAL 1 37
FT CHAIN 38 1135 IGA FC RECEPTOR.
FT PROPEP 1136 1164 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 434 534 IG-LIKE DOMAIN.
FT DOMAIN 199 438 IGA-BINDING (POTENTIAL).
FT DOMAIN 439 826 IGA-BINDING (POTENTIAL).
FT DOMAIN 827 945 PRO-RICH REPEATS.
FT SITE 1132 1136 LPXIG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1135 1135 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1164 AA: 131051 MW: 650894AF720A5474 CRC54:

Query Match 6.5%; Score 83.5; DB 1; Length 1164.
Best Local Similarity 24.5%; Pred. No. 28;
Matches 39; Conservative 24; Mismatches 65; Indels 81; Gaps 7;
QY 73 RYPCMKRHRDFKSAPEGYVQ:ERTIFKDCNYKTRAEVKEFDHIVNA:TELKGIQFK 131
DE 930 KLPDVPKLPDAPK--LPGLNKVGAVPTSDGN--TKVTTFDQKPTDAKLLKLVETK 985
QY 132 EDGNLGHKJ-----SYNNSHVYIMADKQNGIKVNFKIRINIEDG 174
DE 996 ELADKIAHKITGGTVKRVFVLSLSKGGKETHVNGERTVRLALQGTGS---DVHYVVKENG 1042
QY 175 SVOLADHYQONTPLSDGFSVLLPDNHY--LSTQSALSCKDPN 212
DE 1043 DLERIPSKVEN-----QGVVFKTNHFSFAIKTLKSDQN 1076

Search completed: July 16, 2003, 16:11:17
Job time : 13.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Computer 101

one protein - protein search, using sw model

Run on: July 16, 2003, 16:06:07 ; Search time 41.334 seconds
(without alignments)
1186.44. Billion cell updates/sec

Title: US-09-967-301-2

Perfect score: 1275

Sequence: 1 MSKGEELFGVVPILVELDQ.....VILDEVFAGITHDELYK 238

Scoring table: BLOSUM62

Gapop 15.0 , Gapext 3.5

Searched: 671560 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671590

Minimum DB seq length: 6

Maximum DB seq length: 200600000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP_TREMBL_2.1.1*
- 2: sp_archaea*
- 3: sp_bacteria*
- 4: sp_fungi*
- 5: sp_invertebrate*
- 6: sp_mammal*
- 7: sp_plant*
- 8: sp_organalle*
- 9: sp_hago*
- 10: sp_plant*
- 11: sp_robot*
- 12: sp_virus*
- 13: sp_vertebrate*
- 14: sp_unclassified*
- 15: sp_virus*
- 16: sp_bacteriap*
- 17: sp_archaeap*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	Description
1	1271	99.7	238	5	Q27903
2	1264	99.1	238	5	Q33125
3	1229	96.4	238	5	Q17105
4	1214	95.2	238	5	Q17106
5	1211	95.0	758	4	Q96702
6	1108	86.9	238	5	Q8WPT3
7	1102	86.4	238	5	Q8WTC6
8	1094	85.8	238	5	Q8WTC4
9	1092	85.6	238	5	Q8WTC5
10	1091	85.6	238	5	Q8WTC9
11	1090	85.6	238	5	Q8WTC8
12	1089	85.4	238	5	Q8WTC7
13	1087	85.3	238	5	Q8WTC5
14	259.5	20.4	225	5	Q950A7
15	253	19.8	225	5	Q963F5
16	244.5	19.2	266	5	Q906Y3

17	243.5	18.9	236	5	Q8T5C0
18	238	18.7	225	5	Q8T5F1
19	215	16.9	229	5	Q905Y6
20	214.5	16.8	234	5	Q8T5F2
21	213.5	16.7	235	5	Q8T5F0
22	212	16.6	227	5	Q962P9
23	211.5	16.5	225	5	Q905Y8
24	209.5	16.4	238	5	Q9BLV9
25	208	16.3	227	5	Q95VTC
26	205	16.1	232	5	Q905Y7
27	204.5	16.0	232	5	Q905Y5
28	204	16.0	221	5	Q95904
29	203.5	16.0	225	5	Q8T5Y9
30	203.5	15.7	232	5	Q8G228
31	196	15.4	230	5	Q9GTC7
32	183.5	15.2	229	5	Q9GTC6
33	180.5	14.9	233	5	Q963Y9
34	149.5	14.9	238	5	Q9B420
35	187.5	14.7	227	5	Q95885
36	183.5	14.4	231	5	Q906Y5
37	179.5	14.1	227	5	Q95886
38	178	14.0	229	5	Q8T5E7
39	173.5	13.5	227	5	Q95811
40	172	13.3	231	5	Q906Y4
41	172	13.5	234	5	Q8T5F3
42	167.5	13.1	231	5	Q8T404
43	166.5	13.1	231	5	Q8T5E8
44	156	12.2	231	5	Q8T0T8
45	149	11.7	231	5	Q8T5E9

ALIGNMENTS

RESULT 1	Q27903	PRELIMINARY:	PRT:	238 AA
TD	Q27903			
AC	Q27903			
BT	Q1-NOV-1996 (TRENBLrel_01, Created)			
DT	Q1-NOV-1996 (TRENBLrel_01, Last sequence update)			
DT	Q1-JUN-2002 (TRENBLrel_21, Last annotation update)			
DE	Green fluorescent protein:			
GN	GFP			
OS	unidentified.			
SC	unclassified.			
FX	NCBI TaxID: 32644:			
AX	all			
EX	SEQUENCE FROM N.A.			
EX	MEDLINE:9229832; PubMed:9154991;			
KA	Kawakida G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.			
RT	Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage."			
KL	Plant Mol. Biol. 33:985-999(1997).			
EM	EMBL: X96418; CAA65278.1;			
DS	USSP: Q4212; 18P22			
DR	InterPro: IPR000786; Green_fl_protein.			
DR	Pfam: PF01353; GFP.			
DR	PRINTS: PR01229; GFP_QRESQFNI.			
DR	ProDom: P0613756; Green_fl_protein_1.			
SC	SEQUENCE 238 AA: 26884 MW: CA932D47262AF2D3 CRC64:			

Query Match 99.7% Score 1271 DB 5 Length 238
Res. Local Similarity 99.6% Pred. No. 1.2e-97
Matches 237: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MSKGEELFGVVPILVELDQVNGHKFVSSEGGATYKTLKICTGK:FPWPPL 60		
DB	1	MKGEEELFGVVPILVELDQVNGHKFVSSEGGATYKTLKICTGK:FPWPPL 60		
QY	61	VTTESYGVCFSRYPDHMKRRDFKFSAMPEGYVQERTIFFKDGNTKTRAEVKFGDILV 120		
DB	61	VTTESYGVCFSRYPDHMKRRDFKFSAMPEGYVQERTIFFKDGNTKTRAEVKFGDILV 120		

Matches 198; Conservative 19; Mismatches 21; Indels 0; Gaps 0

QY 1 MSKGEELFTGVVPIVELLDGVNGHKSFSVSGESGDATYKLTIKFKTCTTAKLPVPWFTL 60
 DB 1 MSKGEELFTGVVPIVELLDGVNGHKSFSVSGESGDATYKLTIKFKTCTTAKLPVPWFTL 60
 QY 61 VTTSYGVCFSRYVDHMKRHDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 DB 61 VTTSYGVCFSRYVDHMKRHDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 QY 62 VTLSYGICFARYPERHKKNDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 DB 62 VTLSYGICFARYPERHKKNDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 QY 121 NRLEKGDIFKEDGNILGKLEYNHNSVYIMAKKNGKLVNFKHNEISGVSGLAD 180
 DB 121 NRLEKGDIFKEDGNILGKLEYNHNSVYIMAKKNGKLVNFKHNEISGVSGLAD 180
 QY 181 HYCONTDIFGCVLLPDNHVLTSTQATSKDNEKADHNVLLFVTAAGTTHGMFELYK 240
 DB 181 HYCONTDIFGCVLLPDNHVLTSTQATSKDNEKADHNVLLFVTAAGTTHGMFELYK 240
 QY 181 HYCONTDIFGCVLLPDNHVLTSTQATSKDNEKADHNVLLFVTAAGTTHGMFELYK 240
 DB 181 HYCONTDIFGCVLLPDNHVLTSTQATSKDNEKADHNVLLFVTAAGTTHGMFELYK 240

RESULT 9

Q8WTC9 PRELIMINARY: PRT: 238 AA

Query Match 85.6%; Score 1091; DB 5; Length 238;
 Best Local Similarity 82.4%; Pred. No. 9.9e-83;
 Matches 196; Conservative 16; Mismatches 23; Indels 0; Gaps 0

QY 1 MSKGEELFTGVVPIVELLDGVNGHKSFSVSGESGDATYKLTIKFKTCTTAKLPVPWFTL 60
 DB 1 MSKGEELFTGVVPIVELLDGVNGHKSFSVSGESGDATYKLTIKFKTCTTAKLPVPWFTL 60
 QY 61 VTTSYGVCFSRYVDHMKRHDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 DB 61 VTTSYGVCFSRYVDHMKRHDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 QY 62 VTLSYGICFARYPERHKKNDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 DB 62 VTLSYGICFARYPERHKKNDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 QY 121 NRLEKGDIFKEDGNILGKLEYNHNSVYIMAKKNGKLVNFKHNEISGVSGLAD 180
 DB 121 NRLEKGDIFKEDGNILGKLEYNHNSVYIMAKKNGKLVNFKHNEISGVSGLAD 180
 QY 181 HYCONTDIFGCVLLPDNHVLTSTQATSKDNEKADHNVLLFVTAAGTTHGMFELYK 240
 DB 181 HYCONTDIFGCVLLPDNHVLTSTQATSKDNEKADHNVLLFVTAAGTTHGMFELYK 240

RESULT 9

Q8WTC9 PRELIMINARY: PRT: 238 AA

Query Match 85.6%; Score 1091; DB 5; Length 238;
 Best Local Similarity 82.4%; Pred. No. 9.9e-83;
 Matches 196; Conservative 16; Mismatches 23; Indels 0; Gaps 0

01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 Green fluorescent protein.
 GFP.
 Aequorea macrodactyla.
 Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
 Aequoreidae; Aequorea.
 NCBI_TaxID=147615;
 [1]
 SEQUENCE FROM N.A.
 STRAIN:GFPX161;
 Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan P.Q., Chen M.,
 Li S.J., Xia N.S.;
 "Colorful mutants of green fluorescent protein from Aequorea
 macrodactyla";
 Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 EMBL: AF435437; AAL33913; 1;
 InterPro: IPR000786; Green_fl_protein;
 Pfam: PF01353; GFP; 1;
 ProDom: PD013756; Green_fl_protein; 1;
 SEQUENCE 238 AA; 26957 MW; 5F8DA192173CB84D CRC64;

Query Match 85.6%; Score 1092; DB 5; Length 238;
 Best Local Similarity 82.4%; Pred. No. 9.2e-83;
 Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0

QY 1 MSKGEELFTGVVPIVELLDGVNGHKSFSVSGESGDATYKLTIKFKTCTTAKLPVPWFTL 60
 DB 1 MSKGEELFTGVVPIVELLDGVNGHKSFSVSGESGDATYKLTIKFKTCTTAKLPVPWFTL 60
 QY 61 VTTSYGVCFSRYVDHMKRHDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 DB 61 VTTSYGVCFSRYVDHMKRHDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 QY 62 VTLSYGICFARYPERHKKNDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 DB 62 VTLSYGICFARYPERHKKNDFFKSAMHFGYVQERTIEKDDSNYKTRAEVKEFEDTLV 120
 QY 121 NRLEKGDIFKEDGNILGKLEYNHNSVYIMAKKNGKLVNFKHNEISGVSGLAD 180
 DB 121 NRLEKGDIFKEDGNILGKLEYNHNSVYIMAKKNGKLVNFKHNEISGVSGLAD 180
 QY 181 HYCONTDIFGCVLLPDNHVLTSTQATSKDNEKADHNVLLFVTAAGTTHGMFELYK 240
 DB 181 HYCONTDIFGCVLLPDNHVLTSTQATSKDNEKADHNVLLFVTAAGTTHGMFELYK 240

RESULT 10

Q8WTC9 PRELIMINARY: PRT: 238 AA

Query Match 85.6%; Score 1091; DB 5; Length 238;
 Best Local Similarity 82.4%; Pred. No. 9.9e-83;
 Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0

```

QY 1 MSKGEELFTGVPTLVFLVDGDNVNGHKSVSSEGGDAIVYKLTLEKLTCTTCKLFPVWEL 60
DB 1 MSKGEELFTGVPTLVFLVDGDNVNGHKSVSSEGGDAIVYKLTLEKLTCTTCKLFPVWEL 60
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
QY 61 VTTISYGLICFARYPEHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
DB 61 VTTISYGLICFARYPEHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
QY 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFKIRHNIEGVSQVLAQ 180
DB 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFKIRHNIEGVSQVLAQ 180
QY 181 HYQNTTIGGGVLLPNDNYLSTQSALSKDPNKRDNHVLLEVTAACTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDNYLSTQSALSKDPNKRDNHVLLEVTAACTHGMDELYK 240
QY 191 HYQNTVPLDGPVLPINHYLSVQTAISKDRNETRDMVTFEFSACGTHGMDELYK 300
DB 191 HYQNTVPLDGPVLPINHYLSVQTAISKDRNETRDMVTFEFSACGTHGMDELYK 300

RESULT 11
Q8WTC8 PRELIMINARY: PRT: 238 AA.
AC Q8WTC8: 85.4%, Score 1089, DB 5, Length 238,
DT 01-MAR-2002 (TRENBLrel, 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel, 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel, 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae.
CC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen X.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435430; AAL33915;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27002 MW: 355842982640018 CRC64:

Query Match 85.4%, Score 1089, DB 5, Length 238,
Best Local Similarity 82.4%, Prod. No. 1.5e-82,
Matches 196, Conservative 19, Mismatches 23, Indels 0, Gaps 0;

QY 1 MSKGEELFTGVPTLVFLVDGDNVNGHKSVSSEGGDAIVYKLTLEKLTCTTCKLFPVWEL 60
DB 1 MSKGEELFTGVPTLVFLVDGDNVNGHKSVSSEGGDAIVYKLTLEKLTCTTCKLFPVWEL 60
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
QY 61 VTTISYGLICFARYPEHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
DB 61 VTTISYGLICFARYPEHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
QY 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFKIRHNIEGVSQVLAQ 180
DB 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFKIRHNIEGVSQVLAQ 180
QY 181 HYQNTTIGGGVLLPNDNYLSTQSALSKDPNKRDNHVLLEVTAACTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDNYLSTQSALSKDPNKRDNHVLLEVTAACTHGMDELYK 240
QY 191 HYQNTVPLDGPVLPINHYLSVQTAISKDRNETRDMVTFEFSACGTHGMDELYK 300
DB 191 HYQNTVPLDGPVLPINHYLSVQTAISKDRNETRDMVTFEFSACGTHGMDELYK 300

RESULT 12
Q8WTC7 PRELIMINARY: PRT: 238 AA.
AC Q8WTC7: 85.4%, Score 1091, DB 5, Length 238,
DT 01-MAR-2002 (TRENBLrel, 20, Created)
DT 01-MAR-2002 (TRENBLrel, 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel, 21, Last annotation update)
DE Green fluorescent protein.

```

```

GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae.
CC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen X.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435430; AAL33915;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27002 MW: 355842982640018 CRC64:

Query Match 85.4%, Score 1089, DB 5, Length 238,
Best Local Similarity 82.4%, Prod. No. 1.5e-82,
Matches 196, Conservative 19, Mismatches 23, Indels 0, Gaps 0;

QY 1 MSKGEELFTGVPTLVFLVDGDNVNGHKSVSSEGGDAIVYKLTLEKLTCTTCKLFPVWEL 60
DB 1 MSKGEELFTGVPTLVFLVDGDNVNGHKSVSSEGGDAIVYKLTLEKLTCTTCKLFPVWEL 60
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
QY 61 VTTISYGLICFARYPEHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
DB 61 VTTISYGLICFARYPEHMKHGHGPKSAMPGYVYQERTIFPKDQSNYKTRAEVFEGLILV 120
QY 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFKIRHNIEGVSQVLAQ 180
DB 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFKIRHNIEGVSQVLAQ 180
QY 181 HYQNTTIGGGVLLPNDNYLSTQSALSKDPNKRDNHVLLEVTAACTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDNYLSTQSALSKDPNKRDNHVLLEVTAACTHGMDELYK 240
QY 191 HYQNTVPLDGPVLPINHYLSVQTAISKDRNETRDMVTFEFSACGTHGMDELYK 300
DB 191 HYQNTVPLDGPVLPINHYLSVQTAISKDRNETRDMVTFEFSACGTHGMDELYK 300

RESULT 13
Q8WTC5 PRELIMINARY: PRT: 238 AA.
AC Q8WTC5: 85.4%, Score 1091, DB 5, Length 238,
DT 01-MAR-2002 (TRENBLrel, 20, Created)
DT 01-MAR-2002 (TRENBLrel, 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel, 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae.
CC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen X.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435432; AAL33917;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27018 MW: 75521EA5534E573A CRC64:

Query Match 85.4%, Score 1087, DB 5, Length 238,
Best Local Similarity 83.2%, Prod. No. 2.1e-82,
Matches 198, Conservative 16, Mismatches 24, Indels 0, Gaps 0;

QY 1 MSKGEELFTGVPTLVFLVDGDNVNGHKSVSSEGGDAIVYKLTLEKLTCTTCKLFPVWEL 60
DB 1 MSKGEELFTGVPTLVFLVDGDNVNGHKSVSSEGGDAIVYKLTLEKLTCTTCKLFPVWEL 60

```



```

Db      1  MSKSELGTGVVILVELDGVGHKPSVEGEGDADQKGLFIPICITCKLPYPTIL 60
QY      62  VTFSYGVQCCSRYPDINKKDEPKSAMPGYQVETIFKKKIDNNKTRKAVKREIATIV 120
Db      11  111 1111111111111111111111111111111111111111111111111 141
QY      63  VTIISYGLICFARYPHMKNDPKSAMTEGYIQRTIFELDQYKYGKGVKPSQTIIV 120
Db      11  111 1111111111111111111111111111111111111111111111111 141
QY      121  NIEKGIQDIFQEDQNIIGHKLYNNHNYVIMAKOKKQKYNKTRINTEDQSGVIAQ 160
Db      11  1111111111111111111111111111111111111111111111111 181
QY      121  NPRIEKGDPQFQDNLGHKILEYNNHNYVIMFKKKNQKLYNKIRNTEGGSGVIAQ 160
Db      11  1111111111111111111111111111111111111111111111111 181
QY      141  HYQONTIIGDSQFVILIPKHKYLSFQSALSKUPKPKRHKVYLLFVYANSLIHEMDLEIK 238
Db      11  1111111111111111111111111111111111111111111111111 258
QY      141  HYQNTVIGDGPVILIPKHKYLSFQATSKQKNTREMDLEIFFSNVGHMDKLEIK 238
Db      11  1111111111111111111111111111111111111111111111111 258

RESULT 14
C95UA7
AC      C95UA7      PRELIMINARY.      ERT: 205 AA.
DT      01-DEC-2001 (TREMBLrel. 19, Created)
ET      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
SI      01-JUN-2002 (TRFMBLrel. 21, Last annotation update)
DE      Cyan fluorescent protein (fragment)
OS      Montastraea cavetnosa (great star coral).
OC      Eukaryota; Metazoa; Chordata; Anthozoa; Sclerozoa; Scleractinia.
OX      Faviidae; Faviidae; Montastraea.
UN      NCBI_TaxID=63558;
PX      111
SEQUENCE FROM N.A.
AP      Paikowsky P.G., Sun Y.J.
RT      "Montastraea cavetnosa (fluorescent protein)."
RG      Submitted (SEP-2001) to the EMBL/GenBank/DBSJ databases.
DR      ENRL: AY052460; AAC: F7905.1;
DR      InterPro: I2R000236; Green_II_protein.
DR      Pfam: PF01353; GFF: ...
DR      ProDom: PD013556; Green_II_protein: 1.
DR      NON_TER      225      225
SQ      SEQUENCE      225 AA; 25375 MW; 520E2F776D0F4024 C60640

```

```

RESULT 15
Q963F5
ID Q963F5      PRELIMINARY:      ERT:  2.5  AA.
AC Q963F5;
UT C1-DEC-2001 (TREMSELrel_19: Created)
PT C1-DEC-2001 (TREMSELrel_19: Last sequence update)
ET 01-JUN-2002 (TREMSELrel_21: Last annotation update)
DE Green fluorescent protein.
OS Montastraea cavernosa (great star coral).
CC Eukaryota; Metazoa; Cnidaria; Anthozoa; Scleractinia;
CC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63598;
RN      [1]

```

Genome version 3.1.6
Copyright (c) 1995 - 2001, GenPart Inc.

OR protein - protein search, using sw model

Run on: July 15, 2004, 16:11:38 : Search time: 2.3 seconds
Method: all-against-all
574 635 Million cell updates/sec

Title: us-09-967-301-3
Patient source: 1226
Sequence: 1 MSKGEITGVPLVLEL63.....VAGFVAGHIEHDEHLEK 124

Scoring table: RLOSCM02

Gapop 10.0 : Gapext 6.5

Searched: 45169 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451869

Minimum DB seq length: 0

Maximum DB seq length: 200360000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA*

- 1: /cgn2.6/prodata/2/pubaa/US07_NEW_PUB.pep.*
- 2: /cgn2.6/prodata/2/pubaa/PC1_NEW_PUB.pep.*
- 3: /cgn2.6/prodata/2/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2.6/prodata/2/pubaa/US06_PUBCMP.pep.*
- 5: /cgn2.6/prodata/2/pubaa/PC1_US_PUBCMP.pep.*
- 6: /cgn2.6/prodata/2/pubaa/US07_PUBCMP.pep.*
- 7: /cgn2.6/prodata/2/pubaa/US06_NEW_PUB.pep.*
- 8: /cgn2.6/prodata/2/pubaa/US06_PUBCMP.pep.*
- 9: /cgn2.6/prodata/2/pubaa/US05_NEW_PUB.pep.*
- 10: /cgn2.6/prodata/2/pubaa/US05_PUBCMP.pep.*
- 11: /cgn2.6/prodata/2/pubaa/US16_NEW_PUB.pep.*
- 12: /cgn2.6/prodata/2/pubaa/US16_PUBCMP.pep.*
- 13: /cgn2.6/prodata/2/pubaa/US06_NEW_PUB.pep.*
- 14: /cgn2.6/prodata/2/pubaa/US06_PUBCMP.pep.*

Pred. No. is the number of residues predicted by chance to have a score greater than or equal to the score of the result, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Match	Length	DB ID	Description
1	1258	98.6	238	9	US-10-057-505-2
2	1258	98.6	238	10	US-09-967-301-3
3	1258	98.6	238	12	US-10-024-646-2
4	1255	98.4	238	9	US-09-899-954B-2
5	1254	98.3	238	9	US-09-899-954B-2
6	1254	98.3	238	9	US-09-900-345A-126
7	1254	98.3	238	9	US-10-121-259-10
8	1254	98.3	238	9	US-10-221-651-6
9	1254	98.3	238	10	US-09-920-522-4
10	1254	98.3	238	10	US-09-852-030-1
11	1252	98.1	244	9	US-09-899-954B-9
12	1252	98.1	244	9	US-09-899-954B-10
13	1252	98.1	244	9	US-09-899-954B-12
14	1252	98.1	244	9	US-09-899-954B-14
15	1252	98.1	244	9	US-09-899-954B-16
16	1252	98.1	244	9	US-09-899-954B-18
17	1251	98.0	239	9	US-09-887-784-4
18	1251	98.0	243	9	US-09-900-345A-60
19	1251	98.0	243	9	US-09-900-345A-62

1251	98.6	243	9	US-09-900-345A-64	Sequence 64, Appl
1251	98.6	243	9	US-09-900-345A-66	Sequence 66, Appl
1251	98.6	243	9	US-09-900-345A-68	Sequence 68, Appl
1251	98.6	243	9	US-09-900-345A-70	Sequence 70, Appl
1251	98.6	1070	9	US-10-001-486B-2	Sequence 2, Appl
1251	98.6	1099	9	US-10-259-864-4	Sequence 4, Appl
1251	98.6	1147	9	US-10-259-864-1	Sequence 1, Appl
1251	98.6	1452	12	US-10-050-673-2	Sequence 2, Appl
1251	98.6	247	9	US-09-899-954B-36	Sequence 36, Appl
1251	98.6	243	9	US-09-900-345A-54	Sequence 54, Appl
1251	98.6	243	9	US-09-900-345A-56	Sequence 56, Appl
1251	98.6	243	9	US-09-900-345A-58	Sequence 58, Appl
1251	98.6	243	9	US-09-900-345A-114	Sequence 114, Appl
1251	98.6	243	9	US-09-900-345A-116	Sequence 116, Appl
1251	98.6	243	9	US-09-900-345A-128	Sequence 128, Appl
1251	98.6	243	9	US-09-900-345A-120	Sequence 120, Appl
1251	98.6	244	9	US-09-899-954B-6	Sequence 6, Appl
1251	98.6	244	9	US-09-899-954B-22	Sequence 22, Appl
1251	98.6	244	9	US-09-899-954B-22	Sequence 22, Appl
1249	97.9	243	9	US-09-900-345A-4	Sequence 4, Appl
1249	97.9	243	9	US-09-900-345A-6	Sequence 6, Appl
1249	97.9	243	9	US-09-900-345A-8	Sequence 8, Appl
1249	97.9	243	9	US-09-900-345A-10	Sequence 10, Appl
1249	97.9	243	9	US-09-900-345A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-10-057-505-2
? Sequence 2, Application US/10057505
? Patent No. US2002016474A1
? GENERAL INFORMATION:
? APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
? APPLICANT: AURORA BIOSCIENCES CORPORATION
? APPLICANT: TSEH, Roger
? APPLICANT: HEM, Roger
? APPLICANT: CHITT, Andrew
? TITLE OF INVENTION: TANDEM FLUORESCENT PROTEIN CONSTRUCTS
? FILE REFERENCE: REGEN1260-3
? CURRENT APPLICATION NUMBER: US/10057505
? PRIORITY FILING DATE: 2002-11-25
? PRIOR APPLICATION NUMBER: US 08/742,553
? PRIOR FILING DATE: 1997-01-31
? PRIOR APPLICATION NUMBER: US 09/396,003
? PRIOR FILING DATE: 1999-04-13
? NUMBER OF SEQ ID NOS: 32
? SOFTWARE: Patent 16 version 3.1
? SEQ ID NO 2
? LENGTH: 238
? TYPE: PRT
? ORGANISM: Aequorea victoria
US-10-057-505 2

Query Match	98.6%	Score 1258	DB 9	Length 238
Posit Local Similarity	98.3%	Pred. No. 96-115		
Mismatches	235	Conservative	0	Mismatches 3
Indels				Gaps 0
QY	3	MSKGEELTAVPILVELDGVNHSKPSVSGESDATYKLTLPICITKLPVWPPL	60	
DB	1			
QY	61	VTTLSYVQCFSSVPDMKHDFPKSAMPHGYVCEWTIFKODGNYKTRAFKPEECLTV	120	
DB	61	VTFYSVQCFSSVPDMKHDFPKSAMPHGYVCEWTIFKODGNYKTRAFKPEECLTV	120	
QY	121	NRIELKGLDFKESGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRNIEFGVOLAD	180	
DB	121	NRIELKGLDFKESGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRNIEFGVOLAD	180	

OTHER INFORMATION: Green fluorescent protein
US-09-852-036-1

Query Match 98.1%, Score 1254, DP 126, Length 244;
Best Local Similarity 98.1%, Pred. No. 2.4e-114;
Matches 234; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVVELLGGVGNHKKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
DB 1 MSKGEELFTGVVPIVVELLGGVGNHKKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60

QY 61 VTLSYGVOCFSRYPDHKKRHFESKAMPEGVQERTFFKDDGNKKEPAVKKFEGLTV 120
DB 61 VTLSYGVOCFSRYPDHKKRHFESKAMPEGVQERTFFKDDGNKKEPAVKKFEGLTV 120

QY 121 NRLEKGIQDFKEDGNILGHKLEYNNSHNVYIMAKOKNGIKVNEKIRHNIELGVGLAD 180
DB 121 NRLEKGIQDFKEDGNILGHKLEYNNSHNVYIMAKOKNGIKVNEKIRHNIELGVGLAD 180

QY 181 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRKHVYLVGVTAAGLTHGMDELYK 244
DB 181 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRKHVYLVGVTAAGLTHGMDELYK 244

RESULT 11
US-09-899-954B-8
Sequence 9, Application US/09899954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
FILE REFERENCE: 10338-60S
CURRENT APPLICATION NUMBER: US/09/899,954B
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU PPR077
PRIOR FILING DATE: 1999-01-08
PRIOR FILING DATE: PCT/AU00/2607
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patent version 3.1
SEQ ID NO 8
LENGTH: 244
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Gen(CTA)6 GFP construct
US-09-899-954B-8

Query Match 98.1%, Score 1254, DP 9, Length 244;
Best Local Similarity 92.9%, Pred. No. 2.4e-114;
Matches 234; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVVELLGGVGNHKKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
DB 7 LSKGEELFTGVVPIVVELLGGVGNHKKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60

QY 61 VTLSYGVOCFSRYPDHKKRHFESKAMPEGVQERTFFKDDGNKKEPAVKKFEGLTV 120
DB 67 VTLSYGVOCFSRYPDHKKRHFESKAMPEGVQERTFFKDDGNKKEPAVKKFEGLTV 120

QY 121 NRLEKGIQDFKEDGNILGHKLEYNNSHNVYIMAKOKNGIKVNEKIRHNIELGVGLAD 180
DB 127 NRLEKGIQDFKEDGNILGHKLEYNNSHNVYIMAKOKNGIKVNEKIRHNIELGVGLAD 180

QY 181 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRKHVYLVGVTAAGLTHGMDELYK 244
DB 187 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRKHVYLVGVTAAGLTHGMDELYK 244

RESULT 12
US-09-899-954B-10
Sequence 10, Application US/09899954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
FILE REFERENCE: 10338-60S
CURRENT APPLICATION NUMBER: US/09/899,954B
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU PPR077
PRIOR FILING DATE: 1999-01-08
PRIOR FILING DATE: PCT/AU00/2607
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patent version 3.1
SEQ ID NO 10
LENGTH: 244
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Leu(CTC)6 GFP construct
US-09-899-954B-10

Query Match 98.1%, Score 1254, DP 9, Length 244;
Best Local Similarity 97.9%, Pred. No. 2.4e-114;
Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVVELLGGVGNHKKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
DB 7 LSKGEELFTGVVPIVVELLGGVGNHKKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60

QY 61 VTLSYGVOCFSRYPDHKKRHFESKAMPEGVQERTFFKDDGNKKEPAVKKFEGLTV 120
DB 67 VTLSYGVOCFSRYPDHKKRHFESKAMPEGVQERTFFKDDGNKKEPAVKKFEGLTV 120

QY 121 NRLEKGIQDFKEDGNILGHKLEYNNSHNVYIMAKOKNGIKVNEKIRHNIELGVGLAD 180
DB 127 NRLEKGIQDFKEDGNILGHKLEYNNSHNVYIMAKOKNGIKVNEKIRHNIELGVGLAD 180

QY 181 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRKHVYLVGVTAAGLTHGMDELYK 244
DB 187 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRKHVYLVGVTAAGLTHGMDELYK 244

RESULT 13
US-09-899-954B-12
Sequence 12, Application US/09899954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
FILE REFERENCE: 10338-60S
CURRENT APPLICATION NUMBER: US/09/899,954B
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU PPR077
PRIOR FILING DATE: 1999-01-08
PRIOR FILING DATE: PCT/AU00/2607
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patent version 3.1
SEQ ID NO 12
LENGTH: 244
TYPE: PRT
ORGANISM: Artificial Sequence

```

1  FEATURE:
2  OTHER INFORMATION: Leu(CTT)6 GFP construct
3  US-09-899-954B-12
4
5  Query Match          98.1%   Score 1252:   DB 9:   Length 244:
6  Best Local Similarity 97.9%   Pred. No. 2.4e-114:
7  Matches 233:   Conservative 1:   Mismatches 4:   Indels 0:   Gaps 0:
8
9  CY 1 MSKGEELFTGVVPIVLVDGVNGHKSFSVSGEGSDATYGLTKFKICTTGKLPVWPPI 60
10 DB 7 LSKGEELFTGVVPIVLVDGVNGHKSFSVSGEGSDATYGLTKFKICTTGKLPVWPPI 60
11
12 CY 61 VTLSYGVCFSRYPDHMKRHDFEKSAMPEGYVQERTIFFKDGNYKTRAEVKFEGDTLV 120
13 DB 67 VTLSYGVCFSRYPDHMKRHDFEKSAMPEGYVQERTIFFKDGNYKTRAEVKFEGDTLV 120
14
15 CY 121 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
16 DB 127 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
17
18 CY 181 HYQONTPIGDGPVLDPNHVLTQSALSKDPNEKRDHMLVGFVTAATTHOMLELYK 238
19 DB 187 HYQONTPIGDGPVLDPNHVLTQSALSKDPNEKRDHMLVGFVTAATTHOMLELYK 244
20
21 RESULT 14
22 US-09-899-954B-14
23 Sequence 14: Application US/0989954B
24 Patent No. US20020157155A1
25 GENERAL INFORMATION:
26 APPLICANT: The University of Queensland (all designated States except US)
27 APPLICANT: Frazer, Ian Hector (US only)
28 APPLICANT: Zhou, Jian (US only)
29 APPLICANT: Botella, Jose (US only)
30 TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein in
31 TITLE OF INVENTION: Target cell or tissue of a plant
32 FILE REFERENCE: 10338-6J5
33 CURRENT APPLICATION NUMBER: US/09/899-954B
34 PRIOR FILING DATE: 2002-03-15
35 PRIOR APPLICATION NUMBER: AU PP8C77
36 PRIOR FILING DATE: 1999-01-08
37 PRIOR APPLICATION NUMBER: PCT/AU00/0007
38 PRIOR FILING DATE: 2000-01-07
39 NUMBER OF SEQ ID NOS: 38
40 SOFTWARE: PatentIn version 3.1
41 SEQ ID NO 14
42 LENGTH: 244
43 TYPE: PR1
44 ORGANISM: Artificial Sequence
45 FEATURE:
46 OTHER INFORMATION: Leu(TTA)6 GFP construct
47 US-09-899-954B-14
48
49 Query Match          98.1%   Score 1252:   DB 9:   Length 244:
50 Best Local Similarity 97.9%   Pred. No. 2.4e-114:
51 Matches 233:   Conservative 1:   Mismatches 4:   Indels 0:   Gaps 0:
52
53 CY 1 MSKGEELFTGVVPIVLVDGVNGHKSFSVSGEGSDATYGLTKFKICTTGKLPVWPPI 60
54 DB 7 LSKGEELFTGVVPIVLVDGVNGHKSFSVSGEGSDATYGLTKFKICTTGKLPVWPPI 60
55
56 CY 61 VTLSYGVCFSRYPDHMKRHDFEKSAMPEGYVQERTIFFKDGNYKTRAEVKFEGDTLV 120
57 DB 67 VTLSYGVCFSRYPDHMKRHDFEKSAMPEGYVQERTIFFKDGNYKTRAEVKFEGDTLV 120
58
59 CY 121 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
60 DB 127 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
61
62 CY 181 HYQONTPIGDGPVLDPNHVLTQSALSKDPNEKRDHMLVGFVTAATTHOMLELYK 238
63 DB 187 HYQONTPIGDGPVLDPNHVLTQSALSKDPNEKRDHMLVGFVTAATTHOMLELYK 244
64
65 RESULT 15
66 US-09-899-954B-15
67 Sequence 15: Application US/9989954B
68 Patent No. US20020157155A1
69 GENERAL INFORMATION:
70 APPLICANT: The University of Queensland (all designated States except US)
71 APPLICANT: Frazer, Ian Hector (US only)
72 APPLICANT: Zhou, Jian (US only)
73 APPLICANT: Botella, Jose (US only)
74 TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein in
75 TITLE OF INVENTION: Target cell or tissue of a plant
76 FILE REFERENCE: 10338-6J5
77 CURRENT APPLICATION NUMBER: US/09/899-954B
78 PRIOR FILING DATE: 2002-03-15
79 PRIOR APPLICATION NUMBER: AU PP8C77
80 PRIOR FILING DATE: 1999-01-08
81 PRIOR APPLICATION NUMBER: PCT/AU00/0007
82 PRIOR FILING DATE: 2000-01-07
83 NUMBER OF SEQ ID NOS: 38
84 SOFTWARE: PatentIn version 3.1
85 SEQ ID NO 15
86 LENGTH: 244
87 TYPE: PR1
88 ORGANISM: Artificial Sequence
89 FEATURE:
90 OTHER INFORMATION: Leu(CTT)6 GFP construct
91 US-09-899-954B-15
92
93 Query Match          98.1%   Score 1252:   DB 9:   Length 244:
94 Best Local Similarity 97.9%   Pred. No. 2.4e-114:
95 Matches 233:   Conservative 1:   Mismatches 4:   Indels 0:   Gaps 0:
96
97 CY 1 MSKGEELFTGVVPIVLVDGVNGHKSFSVSGEGSDATYGLTKFKICTTGKLPVWPPI 60
98 DB 7 LSKGEELFTGVVPIVLVDGVNGHKSFSVSGEGSDATYGLTKFKICTTGKLPVWPPI 60
99
100 CY 61 VTLSYGVCFSRYPDHMKRHDFEKSAMPEGYVQERTIFFKDGNYKTRAEVKFEGDTLV 120
101 DB 67 VTLSYGVCFSRYPDHMKRHDFEKSAMPEGYVQERTIFFKDGNYKTRAEVKFEGDTLV 120
102
103 CY 121 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
104 DB 127 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
105
106 CY 181 HYQONTPIGDGPVLDPNHVLTQSALSKDPNEKRDHMLVGFVTAATTHOMLELYK 238
107 DB 187 HYQONTPIGDGPVLDPNHVLTQSALSKDPNEKRDHMLVGFVTAATTHOMLELYK 244
108

```

Search completed: July 16, 2003, 16:30:10
CPU time : 29 secs

Genome version 5.1.4
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2003 15:57:12 : Search time 40.672 seconds
(without alignment)
100.000 hits per cell updates/sec

Title: US-09-967-301-3
Perfect score: 1276
Sequence: 1 KSKKRLATGVVILVERDG.....VLIDPVAGDILHEDMKR LK

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 5616422 residues

Total number of hits satisfying chosen parameters: 17224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match OK
Maximum Match 100%
Listing first 45 summaries

Database : PIR73:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARY

Result	Query	Match	Length	DB	ID	Accession
1	1248	97.8	248	1	Q05514	green-fluorescent protein
2	107	8.4	785	2	H72428	hypothetical protein
3	99.5	7.8	887	2	B85592	hypothetical protein
4	92.5	7.2	534	1	NICLHA	nitrogenase (EC 1.1.1.1)
5	90	7.1	1224	1	FRH248	nitrogenase (EC 1.1.1.1)
6	89.5	7.0	861	2	H64102	nitrogenase (EC 1.1.1.1)
7	88.5	6.9	370	2	E70490	nitrogenase (EC 1.1.1.1)
8	86	6.9	353	2	B84441	nitrogenase (EC 1.1.1.1)
9	87.5	6.9	792	2	Q04578	nitrogenase (EC 1.1.1.1)
10	87.5	6.9	804	2	E64102	nitrogenase (EC 1.1.1.1)
11	87.5	6.9	222	1	A56228	DNA-directed RNA polymerase
12	87	6.8	357	2	G61355	DNA-directed RNA polymerase
13	87	6.8	432	2	G61314	DNA-directed RNA polymerase
14	86.5	6.8	263	2	S53488	DNA-directed RNA polymerase
15	86.5	6.8	860	2	AC5582	DNA-directed RNA polymerase
16	86.5	6.8	253	2	E75414	DNA-directed RNA polymerase
17	85.5	6.7	653	2	H83017	DNA-directed RNA polymerase
18	85	6.7	612	2	T06595	DNA-directed RNA polymerase
19	84.5	6.4	730	1	C14932	DNA-directed RNA polymerase
20	84.5	6.6	889	2	JC5576	hypothetical protein
21	84	6.6	281	2	AC2052	hypothetical protein
22	84	6.6	51	2	Q95338	hypothetical protein
23	84	6.6	874	2	Q04360	hypothetical protein
24	83.5	6.5	613	2	A99552	oligodeoxyribonucleotidyl transferase
25	83	6.5	578	1	L40734	oligodeoxyribonucleotidyl transferase
26	82.5	6.5	312	2	C61710	oligodeoxyribonucleotidyl transferase
27	82.5	6.5	422	2	B90185	oligodeoxyribonucleotidyl transferase
28	82.5	6.5	502	2	B90183	oligodeoxyribonucleotidyl transferase
29	82.5	6.5	941	2	S29443	oligodeoxyribonucleotidyl transferase

ALIGNMENTS

RESULT 1

JO1514
green-fluorescent protein [validated] - Hydrocodusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence, revision: 03-Dec-1999 #text change 23-Mar-2001
C:Accession: J01514; P03355; S48693; S51330; S51331
R:Flasher, D.C.; Eckhardt, V.K.; Ward, W.W.; Prendergast, F.G.; Gorman, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JO1514; MUID:92175527; PMID:1347277
A:Accession: J01514
A:Molecule type: DNA
A:Residues: 1-107, 'S', 139-232 <PRAL>
A:Cross-references: GR:M62654; NID:q155663; PIRN:AAA27722.1; PIR:q155663
A:Accession: Q01514
A:Molecule type: mRNA
A:Residues: 1-99, 'F', 101-140, 'L', 142-238, 'V', 220-238 <PRA2>
A:Cross-references: GR:M62653; NID:q155660; PIRN:AAA27721.1; PIR:q155661
A:Accession: Q0335
A:Molecule type: protein
A:Residues: 46-64, 77-122, 132-151, 154-183, 185-200 <PRA3>
A:Accession: S48693; NID:q155662; PIRN:AAA27722.1; PIR:q155662
A:Title: Evidence for: redox forms of the Aequorea victoria green fluorescent protein.
A:Reference number: S48693; MUID:94364479; PMID:8582767
A:Accession: S48693
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-26, 'Q', 28-156, 'P', 158-172, 'K', 173-238 <PRA4>
A:Cross-references: GR:L29345; NID:q155664; PIRN:AAA6245.1; PIR:q155664
A:Accession: S48693; NID:q155664; PIRN:AAA6245.1; PIR:q155664
Submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13, 'A', 15-24, 'Q', 26-44, 'K', 46-153, 'S', 155-156, 'P', 158-172, 'K', 173-227, 'V'
A:Cross-references: FMB:YK3359; NID:q155665; PIRN:CAA58769.1; PIR:q155665
A:Accession: S51331
A:Molecule type: protein
A:Residues: 1-24, 'Q', 26-29, 'P', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-172, 'K', 173-208, 'V'
A:Cross-references: FMB:YK3359; NID:q155665; PIRN:CAA58769.1; PIR:q155665
A:Experimental source: clone gfp2
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A58992; PDB:13FL
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'B', 15
A:Note: engineered sequence based on JO1514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1245-1251, 1996
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58993; MUID:98294543; PMID:9631067

IgA B2 receptor pr
IgA B2 receptor pr
hypothetical prote
alpha-amylase homo
lipoxigenase (EC 1
synthetohemotropi
hypothetical prote
ABC type transport
DNA-directed DNA p
hypothetical prote
protein WGG11.7 1
photosystem II chl
proteobactera myosin be
NADH dehydrogenase
conserved tyrosine
trypsin-like phospho

A:Contents: annotation; X-ray crystallography, 1.5 Å structure
C:Comment: This protein is excited by the photochemical reaction (see PIR-A0109V) emitting
C:Comment: The chromophore of this protein is formed by modification of Ser-46hydro-Tyr-
C:Genetics:
A:Gene: GFP
A:Intros: 69/3; 167/3
C:Superfamily: green-fluorescent protein
C:Keywords: Chromoprotein; luminescence
F:65/67/Cross-link: 5-imidazolincine (Ser-Gly) *status experimental;
F:66/Modified site: dehydrotyrosine (Tyr) *status experimental

Query Match 97.8%; Score 124%; DR 1; Length 236;
Best Local Similarity 97.1%; Pred. No. 4; 1e-99;
Matches 231; Conservative 4; Mismatches 3; Gaps 0;
QY 1 MSKGEELFTGVPVILVELDGVNGHKSVSGBGHDATYKGLTKFTCTCTCKLFPVWPPTL 50
DB 1 MSKGEELFTGVPVILVELDGVNGHKSVSGBGHDATYKGLTKFTCTCTCKLFPVWPPTL 50
QY 61 VTLSYGVQCFSRYPDHRKRDHDFKSNMPGVGERTIFPKKDEKYNKTRAEVNEEGTTLV 120
DB 61 VTLSYGVQCFSRYPDHRKRDHDFKSNMPGVGERTIFPKKDEKYNKTRAEVNEEGTTLV 120
QY 121 NRIELKGDREKDNILGKRGYWNHSHNYIMADKQVNAIKVNEKIKENNEAGVGLAD 160
DB 121 NRIELKGDREKDNILGKRGYWNHSHNYIMADKQVNAIKVNEKIKENNEAGVGLAD 160
QY 161 HYQONTFAGCEVLLPQNYSTQSALSKDGNKRDHMLGVTANGLTKNGCELYK 236
DB 161 HYQONTFAGCEVLLPQNYSTQSALSKDGNKRDHMLGVTANGLTKNGCELYK 236

RESULT 2
H72228
Hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999; #text_change 21-Jul-2000
C:Accession: H72228
R:Neisoon, K.E.; Clayton, R.A.; Gill, S.E.; Gellum, M.J.; Pederson, R.L.; Hall, D.H.; Black
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, K.S.; Phillips, G.A.; Richardson, L.J.
C.M.
Nature 356, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from a genome-sequenced organism
A:Reference number: A72260; MUID:94287A16; #text_change 21-Jul-2000
A:Accession: H72228
A:Status: preliminary
A:Molecule type: tRNA
A:Residues: 1745 GARS
A:Cross-references: GB:AB001846; GB:AB000012; NID:41344594; EMBL:AA14594.1; F000449214
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1624

Query Match 9.4%; Score 107; DR 2; Length 761;
Best Local Similarity 19.7%; Pred. No. 9; 4e-33;
Matches 46; Conservative 33; Mismatches 79; Gaps 7;
QY 2 SKSGELFTGVPVILVELDGVNGHKSVSGBGHDATYKGLTKFTCTCTCKLFPVWPPTL 61
DB 15 NKRGSFSTFPGVQAD 46
QY 62 TTLSYGVQCFSRYPDHRKRDHDFKSNMPGVGERTIFPKKDEKYNKTRAEVNEEGTTLV 121
DB 62 TTLSYGVQCFSRYPDHRKRDHDFKSNMPGVGERTIFPKKDEKYNKTRAEVNEEGTTLV 121
QY 122 RTELKGDREKDNILGKRGYWNHSHNYIMADKQVNAIKVNEKIKENNEAGVGLAD 161
DB 89 DYVNGVYL---GSTEDMEIEFPDVTNVL-----KENNELKLYIK-----SIRVDPK 194
QY 182 YQONTFIDGCVLLPQNYSTQSALSKDGNKRDHMLGVTANGLTKNGCELYK 234
DB 135 IRLNYGVVLGSP-----FEP-----IPGVIRKKNYSYGVK 163

RESULT 3

E82590
lucyl-tRNA synthetase X82175 (imported); - Xylella fastidiosa (strain 9450)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000; #text_change 02-Sep-2000
C:Accession: E82590
R:Anonymous; The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59428 below
A:Accession: E82590
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-887 <SIN>
A:Cross-references: GB:AE004031; GB:AE003849; NID:93107309; PIR:AAFE4975.1; GSPDB:GN
R:Simpson, A.C.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Aceneto, M.; Alvarenga, R.
B:Simoes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carriago, D.M.; Carrer
as-Neto, P.; Docena, C.; El-Dorri, H.; Fancinani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Perro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Pr
J.C.; Chuqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, C.B.; Krumpholtz, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.P.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miranda, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,
Rouiques, V.; Rosa, A.C.G.M.; de Rosa Jr., V.E.; de Sa, R.G.; Saitelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.
A:Reference number: A59329
A:Contents: annotation
C:Genetics:
A:Gene: XP2176
C:Superfamily: Leucine-tRNA ligase

Query Match 7.8%; Score 99.5; DR 2; Length 887;
Best Local Similarity 23.2%; Pred. No. 2; 3;
Matches 46; Conservative 28; Mismatches 67; Gaps 10;
QY 45 ITGKLPVPWPTLVTITISYGVQCFSRYPDHRKRDHDFKSNMPGVGERTIFPKKDGNY-- 106
DB 323 TNEQLPV-WVAVFLMAYVTGAVNAVPGEDGKQDF--ANKYGLPIK-VZALKZIKNDF 345
QY 327 -----KTRAEVNEEGTTLVNRLEKGLDKKESGNTLGHKLEYNHNNYV 152
DB 327 -----KTRAEVNEEGTTLVNRLEKGLDKKESGNTLGHKLEYNHNNYV 152
QY 329 STWEDVNRKDYADKTR--EPE---LINSAEFDGQDAFEVLAASPE----- 429
DB 329 STWEDVNRKDYADKTR--EPE---LINSAEFDGQDAFEVLAASPE----- 429
QY 143 MAKKCKNG-IKYNFKIKHNIIDGGVQLADHYQONTPI-----GGDGVLLPDN 198
DB 430 ---RQKGGKRRYNYR-----FWVSRKRRYNYR-----GGDGVLLPDN 198
QY 199 -HYLTQSALSKDGNKTR 215
DB 483 VAFSGTGSPIKTDPSKKK 500

RESULT 4
N1C1MA
nitrogenase (EC 1.18.6.1) molybdenum-iron protein alpha chain (validated) - Clostridi
N:Alternate names: dinitrogenase alpha chain; nitrogenase component 1 alpha chain
C:Species: Clostridium pasteurianum
C:Date: 01-Sep-1981 #sequence_revision 31-Dec-1991; #text_change 19-Jan-2001
C:Accession: S07389; 140816; A00545
R:Wang, S.Z.; Chen, J.S.; Johnson, J.L.
Nucleic Acids Res. 15, 3935, 1987
A:Title: Nucleotide and deduced amino acid sequences of nifD encoding the alpha-subun
A:Reference number: S07389; MUID:87231095; PMID:3473447
A:Accession: S07389
A:Molecule type: DNA
A:Residues: 1-534 <WAN>
A:Cross-references: EMBL:Y00155; NID:940594; PIR:CAA68449.1; PIR:954895
R:Chen, K.C.

A:Title: HEP-CGF, a novel human gene whose product is highly homologous to the alpha
A:Reference number: J04568; MOLID:96194806; PMID:8647451
A:Accession: J04568
A:Molecule type: mRNA
A:Residues: 1-224 <CHO>
A:Cross-references: G3:U2105; NID:q1536573; PID:AR070979.1; PID:q1522665
A:Experimental source: HepG2 hepatocellular carcinoma cell
A:Feature: G.E.: Hanscher, G.; Kusilek, R.; Meyer, H.E.; Metzger, J.W.
A:Feat. Chem: 267, 22305 22309, 1982
A:Title: Identification of xenin, a xenopsin-related peptide, in the human gastric mu
A:Reference number: A44317; MCID:9454515; PMID:1429581
A:Accession: A44317
A:Molecule type: protein
A:Residues: 1-25 <PEP>
A:Experimental source: gastric mucosa
A:Note: sequence extracted from NCBI; backbone (NCBIP:117013)
A:Note: plasma levels of xenin 25, as determined by immunoassay, rise after meals and
C:Genetics:
A:Gene: GJB:GCPA: HEP-CGP
A:Cross-references: G06:4562787; OMIM:601924
A:Map position: 1q23-1q25
C:Superfamily: yeast coatomer complex alpha chain; WD repeat homology
C:Keywords: duplication; hormone; plasma; stomach
F:1-25/Product: xenin; 25 #status experimental <XNF>
F:5-38/Domain: WD repeat homology <WD1>
F:17-25/Product: xenopsin-related peptide #status predicted <XNF>
F:47-80/Domain: WD repeat homology <WD2>
F:89-122/Domain: WD repeat homology <WD3>
F:131-164/Domain: WD repeat homology <WD4>
F:201-234/Domain: WD repeat homology <WD5>
F:245-278/Domain: WD repeat homology <WD6>

	Query Match	7.1%	Score 90:	DB 1:	Length 1224:
	Best Local Similarity	25.9%	Pred. No. 23:		
	Matches	41:	Conservative	26:	Mismatches 59: Indexes 32: Gaps 7:
QY	90	EGYQVETIFFKDDGNYKTRAEVFEGETIVARIE-LKGIDEFKEDGNILKHKLEVYNVSH 148			
DB	632	KGYFEVALLHFVDE---KIRFSLALEGSGNIEIALFAAKALDQKNWEKLGVAL--LGNHQ 688			
QY	149	NYVIMADKQKNIKYNF-----KIRNIEDGQVQ--LADHYQNTPTETGSPVLLPD 197			
DB	689	IVEMCYQR-KNFKDKVSFLYLTGNLEKJRKMKKTAIRKDMGSHYCNALYLPD----- 742			
QY	198	NHVLSTQSAISKDEKEKEDKIVLQFVTRAGLIHQMCE 235			
DB	742	---VSSRVRILNCGSKS-----LAVETAA--THGLDE 769			

RESULT #
H64102
A:Title: Leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)
A:Alternate names: leucyl-tRNA synthetase
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 16-Aug-1995 #ext_change 03-Jun-2002
C:Accession: H64102
A:Residues: 1-361 <TIG>
A:Experimental source: R.D.: Adams, M.D.; White, C.; Clayton, R.A.; Kirkness, E.F.; Kerlavac
A:Feat. Gen: J.B.: Scott, J.; Shirley, R.; Liu, J.; Glodok, A.; Kelley, J.M.; Weidman
A:Feat. Chem: 267, 22305 22309, 1982
A:Feat. Gen: J.B.: Bracken, R.; Fine, L.B.; Pritchman, J.L.; Fuhrmann, J.L.; Geophaen, N.S.M.
A:Feat. Chem: 259, 496-512, 1995
A:Accession: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.G.; Verte
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64600; MOLID:95350630; PMID:7542860
A:Accession: H64102
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown; translation not shown
A:Residues: 1-361 <TIG>
A:Cross-references: GB:U02774; GB:L42023; NID:q1573942; PID:AA02255el.1; PID:q1573943
C:Genetics:
A:Gene: leus
C:Superfamily: leucine-tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match: 7.0%; Score 89.5; DB 2; Length 561;
Best Local Similarity 24.3%; Pred. No. 16;
Matches 45; Conservative 26; Mismatches 67; Indels 47; Gaps 8;
QY 49 TTGKLPVPTLVITLSGVGQCFSSYPDHRKRDPEKSNMPGEGYVQERTFFPKDGNKYK 108
DB 374 TCDKLPIT-KVANFVLMHYGTVGAVMAVPAHQDQ--FFRQKYSLFKQV-APLAHLEID 370
QY 109 KAEVKEGDTLVNFKLGIDFKEDGNILGHKLEYNINSHVIMAK QKNGT--KVN 154
DB 371 TKQAFVERGKLVNDEFQKRF--DGAFTNG-----TADKLEKLVGKQGVN 414
QY 165 FKIRH-----NIEDGVGLADHYQGNFEGGCVLLPNDHYL-STQSA 207
DB 415 YRLDVGWSRORYGAPIMCLLENGDVPA-----FMEDPILPREDVYMDVASPI 467
QY 208 SKDPN 212
DB 466 NADPN 472
RESULT 7
E70360
Iron-sulfur cofactor synthesis protein nifs - Aquifex aeolicus
N:Contains: L-cysteine sulfoxidtransferase (EC 2.8.1.1)
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #taxi_source 07-Dec-1993
C:Accession: E70360
P:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lopez, A.L.; Graham, D.E.; V.
Nature 392, 354-359, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference numbers: A70360; KM19:98194566; PMID:9543329
A:Accession: E70360
A:Status: preliminary; nucleic acid sequence by shotgun fragmentation not shown
A:Molecule type: DNA
A:Residues: 1-370 <SIC>
A:Cross-references: GB:A2000720; NID:42983529; PIDN:AAC07-11-1; PLOC244854; GB:A200075
A:Experimental source: strain V55
C:Genetics:
C:Superfamily: nifs1
C:Keywords: nitrogen fixation protein nifs
F195/Binding site: pyridoxal phosphate sulfoxidtransferase
F318/Active site: Cys (cysteine persulfide intermediate) #status predicted
Query Match: 6.9%; Score 84.5; DB 2; Length 370;
Best Local Similarity 25.4%; Pred. No. 64;
Matches 49; Conservative 30; Mismatches 91; Indels 23; Gaps 8;
QY 3 KGEELFTGW----PILVELD---GVNKHKE-SVSGES--DEENATCKLTAEFCT 49
DB 104 KQVLLIDAVNAQKPIEKANISYATSGKPKPAKSGFLYISPAEYELIVGAEQ 223
QY 50 TKIP-----VPMFTLVITLSGVGQCFSSYPDHRK RDPEKSNMPGEGYVQERTFFPK 109
DB 224 NGKRSCTENWVGLLSAKALETIVSNFSKYQEQKLRDGFENLELEA-DEACIVGKQ 282
QY 104 GNYKTRAEV--KRGDTLVNFKLGIDFKEDGNILGHKLEYNINSHVIMAKKNG 154
DB 283 ERSPISSIVMPKFAKIVNKLSPKGYCSTGCTACISGSEYEPKHKMLKMFSGEKALAM 342
QY 161 IKVNEKIRHNF 173
DB 343 VRFSGLLNKEE 355
RESULT 8
E84941
imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) [unported] - Buchnera sp. (strain
E84941
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #taxi_source 23-Mar-2001
C:Accession: E84941

R:Shienobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 61-66, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.
A:Reference number: A84940; MUID:20445173; PMID:10993077
A:Accession: E84941
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-353 <SIC>
A:Cross-references: GB:AF000399; GSPDB:GN02144
A:Experimental source: strain APS
C:Genetics:
A:Gene: hlsB; B0102
C:Superfamily: hlsB bifunctional enzyme; imidazoleglycerol-phosphate dehydratase homo
C:Keywords: carbon-oxygen lyase; hydro-lyase
Query Match: 6.9%; Score 88; DB 2; Length 353;
Best Local Similarity 24.3%; Pred. No. 65;
Matches 49; Conservative 33; Mismatches 78; Indels 42; Gaps 10;
QY 63 TLSYGVQCFSSYPDHRKRDPEKSNMPGEGYVQERTIF---FKDGNKYKTRAEVKE----- 114
DB 62 TFSFPLQDPS--TADLFMLSVFES---EGVIFDQELCPHFIDRUCVCKPKIKKITEPWL 115
QY 115 -----EKDT---LVNFKLGIDFKEDGNILGHKLEYNINSHVIMAKKNG 159
DB 116 DKLILKSSVICDRDTDMQSNLAKIKCKYKEDCNMLHITKY-LKHNYAHLIETK 174
QY 160 GYKVPKTEHNH-----DGVGLADHYQGNFEGGCVLLPNDHYLSTQSAKDFNFK 214
DB 175 EIKSAIKVLMLEETSKIDGVKFFDMLHLEQISVSSGCM-----NISQGLDLDIUHHT 229
QY 215 -RDMVVLQ--EVDAAGITGCM 233
DB 280 IETGIVLGEALQALGKNG 251
RESULT 9
JC4078
Protective surface antigen D-15 precursor - Haemophilus influenzae (type D)
C:Species: Haemophilus influenzae
A:Variety: type D
C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #taxi_source 29-Sep-1999
C:Accession: JC4078
K:Black, P.S.; Loosmore, S.; Chong, P.; Thomas, W.K.
Gene 156, 97-99, 1995
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus
A:Reference number: JC4078; MUID:95255876; PMID:7777523
A:Accession: JC4078
A:Molecule type: DNA
A:Residues: 1-797 <SIC>
A:Cross-references: GB:U13361; NID:3537447; PIDN:AA85645.1; PID:353744b
A:Experimental source: type b
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
F119/Domains: signal sequence #status predicted <SIC>
F123-797/Product: protective surface antigen D-15 #status predicted <SIC>
Query Match: 6.9%; Score 67.5; DB 2; Length 797;
Best Local Similarity 22.4%; Pred. No. 21;
Matches 49; Conservative 34; Mismatches 73; Indels 54; Gaps 12;
QY 64 TSYGVQCFSSYPDHRKRDPEKSNMPGEGYVQERTIF---FKSAMIEGYVQF-----PTI 98
DB 427 TGYGTESGISYGNASVKQCNFLDTGAASVATCKNDYGVISVNLGYTEPYFKGVSLGGNV 486
QY 39 FPKDGNKYKTRAEVKEGDTLVNFKLGIDFKEDGNILGHKLEYNINSHVIMAKKNG 154
DB 487 FFEVDSKSDTSSTSNKYKRTIYGSNVIL-CFVFNNSYVVGVLGTYNKLINPALEYN--- 542
QY 148 HNYIMADKQK-NGKVN-----FKIRH-----IEDGGVGLADHYQGNFEGGCVLL 195
DB 543 RNLVYOSMKFKGNGIKTINDFDSFGWNNYNSLNRGYEPTKGVKAS-----LG-GRVTI 593

DQ 74 LFADEKICAHNPRLLEYLRQDNKLEKLE-----+GVEELTIKQV--LSILLIYKKELE 125
 QY 173 D 173
 DB 126 D 126
 RESULT 13
 JH0414
 synaptotagmin o-p65-B - electric ray (Discothyrea cmmata)
 N:Alternate names: synaptic vesicle protein o-p65-B
 C:Species: Discothyrea cmmata
 C:Date: 15-Jan-1993 *sequence_revision 15-Jan-1993 *text_change 12-Aug-1999
 C:Accession: JH0414; F50223
 R:Wardland, B.; Miller, K.G.; Schilling, J.; Scheller, P.H.
 Neuron 6, 993-1007, 1991
 A:Title: Differential expression of the p65 gene family.
 A:Reference number: JH0414; MCID:91273991; PMID:1204159
 A:Accession: JH0414
 A:Molecule type: tRNA
 A:Residues: 1-439 <GEN>
 A:Cross-references: GI:864274; NID:0213116; PDB:1AAA; 1225.1; EMBL:023111
 A:Experimental source: Electric organ
 A:Accession: F50223
 A:Molecule type: protein
 A:Residues: MIV, 26-34; XX, 194-199; XX, 201-204; XX, 432-432; D, 434-437 <GEN>
 C:Superfamily: synaptotagmin; protein kinase C C2 region homolog
 C:Keywords: 4; cytogreen; membrane protein; synaptic vesicle
 F:75-101/Domain: hydrophobic <HYD>
 F:153-266/Domain: protein kinase C C2 region homology <K2A>
 F:284-399/Domain: protein kinase C C2 region homology <K2B>
 F:6.46/Binding site: carbohydrate (Asn) (covalent) *status predicted
 Query Match 6.8% Score 97.35 21 21 Length 439
 Best Local Similarity 23.1% Pred. NC. 11
 Matches 59; Conservative 44; Mismatches 32; Indels 94; Gaps 12
 QY 15 LVELEGVNGHKESVSGEMEDATYKLTETETKILIVFWPTIVTIVSYV ---- 64
 DB 38 MNPIDTQNS*EAGVTPGKND-VFELKKEFMELKIDLPKMIYALATVSEGLIT 64
 QY 69 -----QFSPYIPDHKHLFFKAMPDGVCEPTIFPKDGHKTPAFKPS 116
 DB 97 ELGCKAKCKKKKKAKKFKGKK-----KKFKNGKK-DKVSQGN-AGG 148
 QY 117 --DTLVNRIELKGIDFKKGNII--LGEKGVNWNH----- 148
 DB 139 DAETLTSPYKESKAEKELKGLKQSLIVPQKALITVLAALPALDQKALIV 148
 QY 149 -NVYIMADQKN-GLKVN-----FETAHNTEISVQLA-----THVYQNT 147
 DB 199 VVVELLPQKKKKYETKVQKKLTNPTFESVFKVYVQ-ELANKTDMNAVYDF-RESK 257
 QY 168 LGDGVLLPQ-----NHLSTQSALSKDPEKPKDENVLIGFVTA 226
 DB 258 LGDGVVLMKRVLDAGGLDFEWELESAAKEEPKGLGCTSLRVPTAS 305
 RESULT 14
 S53488
 water-stress-inducible protein p52 - chaco potato
 C:Species: Solanum tuberosum (Chaco potato)
 C:Date: 14-Jul-1995 *sequence_revision 21-Jul-1995 *text_change 31-Jul-2000
 C:Accession: S53488
 R:Silhavy, D.; Hutvagner, G.; Barta, E.; Ranfalvi, Z.
 Plant Mol. Biol. 27, 587-595, 1995
 A:Title: Isolation and characterization of a water-stress-inducible cDNA clone from Solanum tuberosum.
 A:Reference number: S53488; MCID:9520251; PMID:7849621
 A:Accession: S53488
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-263 <SIL>
 A:Cross-references: GB:U12439; NID:3607904; PDB:1AAA805.1; PID:36.7905

Query Match 6.8% Score 96.51 21 21 Length 263
 Best Local Similarity 19.8% Pred. NC. 67
 Matches 50; Conservative 14; Mismatches 89; Indels 99; Gaps 9
 QY 31 GSGEGDATYKGLTLKPFICTICKLPVPWPPTIVTILSYGVGCFSEKYPDHHKHEDFKSA 90
 DB 31 GSEKTSIYGEKI-----SYGDDIYG----- 51
 QY 91 GYVQERTFFKDDGNYKTRAF-----VKFEGDTLVNRIEL--KGIDFK----- 131
 DB 52 ---EXTTFDDGNKYGEKTSYGDDTYGKPISYGGONTYGEKTSYKGNKNGYKTSY 167
 QY 132 --EDONILGKHLEY--NYSNNVYIMADCKNGIKVNFALRHNIEDSGVQLAHYQONT 186
 DB 106 GSGDDNKYGEKTSYGVGSGYGEKPKSYGGDDNKYKTSYNEEGYGGVGETNYFENE 167
 QY 187 -----PIGQPVLLPDNHYLSTQSALSKDPEKPKDENVLIGFVTA 220
 DB 158 SETKTSSEYKSEKKHKKHLEHICGLGAAVAAAFALHEKH-----KAEEKPERAKHKKIE 221
 QY 221 LGFVTAAGTTHS 232
 DB 222 EGIAAANAAG 233
 RESULT 15
 A00592
 leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typhi
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: This species has also been called Salmonella typhi.
 C:Date: 09-Nov-2001 *sequence_revision 09-Nov-2001 *text_change 27-Nov-2001
 C:Accession: A00592
 R:Parkhill, G.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
 et al.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
 et al.; Mouton, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, C.; Stevens,
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
 A:Reference number: A00592; PMID:11677608
 A:Accession: A00592
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-850 <PAR>
 A:Cross-references: GB:1A513242; PDB:1CAG; 5125.1; EMBL:16521899; S53488; JH04174
 C:Keywords:
 A:Gene: STY699
 C:Superfamily: leucyl-tRNA lyase

Query Match 6.8% Score 96.51 21 21 Length 869
 Best Local Similarity 23.7% Pred. NC. 28
 Matches 42; Conservative 20; Mismatches 82; Indels 44; Gaps 6
 QY 49 TCGKLPVPWPPTIVTILSYGVGCFSEKYPDHHKHEDFKSA 108
 DB 324 TSEETPV-WANPVMYEGTGVAMVAPGDQD--YEFASKYGLTIKPVILAAQSEEDI 370
 QY 103 KAEVPEGLTIVNRLEKSTLTKEDGNLGEKLEYNNHNVYIMAEKKNKIKVNFKIR 168
 DB 321 SEQALTEKQVLEKSGEFGVLAFAAFNAIAKKI-----AEKSGVGRKAVNPLR 418
 QY 169 HNIENGUV-----LADHYQNTPTGSPVLLPDNHYLSTQSALSKDP 211
 DB 419 ---DWGVSROPYWGAPINVTLEDGTVLPDQLPVLDPEDVVMGDSITSPIKADP 471

Search completed: July 16, 2003, 16:14:39
 Job time : 22.6667 secs

Genome version 5.1.0
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:01:23 : Search time 15.667 seconds
(without alignments)
325.431 Million words searched/sec

Title: US-09-967-301-3
Perfect score: 1276
Sequence: MSKGFELTRVPLVLELGG.....VLNAYVANTHWELKELK

Scoring table: BLSDM62

Gapop 10.0 : Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 250000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt40.*

Freq. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1254	98.3	238	1	GFP_AEQV1	P42212	Aequorea vi
2	99.5	7.6	879	1	SYL_XTEFA	Q98646	Xyrella fas
3	92.5	7.2	533	1	NIPD_GLOPA	P61667	Clostridium
4	90	7.1	1224	1	COPA_HUMAN	P53621	Human sup
5	89.5	7.0	861	1	SYL_HAEIN	P43627	Haemophilus
6	88	6.9	353	1	R157_HYCAI	P32894	Bacteroides
7	87.5	6.8	795	1	D152_HAEIN	P44935	Haemophilus
8	87.5	6.8	797	1	D152_HAEIN	P44934	Haemophilus
9	87.5	6.9	1224	1	COPA_HUMAN	P53621	Human sup
10	87.5	6.9	2222	1	DEAE_YEAST	P24951	Saccharomy
11	87	6.8	357	1	TRNA_GAMTF	Q66982	Camphylid
12	87	6.8	439	1	SYN2_DLS39	P24566	Escherichia
13	86.5	6.8	793	1	D152_HAEIN	P44935	Haemophilus
14	84.5	6.6	700	1	CAN2_HUMAN	P17656	Human sap
15	84.5	6.6	886	1	TH13_HESAD	P97280	Desulfotol
16	84	6.6	374	1	SAAP_BACLO	P44952	Bacillus st
17	83.5	6.5	613	1	PRFP_XYCPU	Q98690	Myxylasma
18	82.5	6.5	912	1	TRXB_HYDRO	P29477	Hydrophila
19	82.5	6.5	824	1	CAN2_PIG	P43187	Porcine scro
20	82.5	6.5	422	1	CAN2_PAS17	P55914	Paracanthas
21	82.5	6.5	659	1	ADAN_HUMAN	Q00116	Human sap
22	82.5	6.5	941	1	GDA_HAT56	P24424	Haemophilus
23	82.5	6.5	144	1	RAL1_STRAP	P27351	Streptococ
24	82	6.4	774	1	ANV2_SCHRC	P24919	Schistosom
25	81.5	6.4	533	1	CP51_CANSA	P53969	Canis lupus
26	81.5	6.4	538	1	GRDF_RAT	P57900	Rattus norv
27	81.5	6.4	798	1	UKA_LRPBE	P13946	Herpes labi
28	81	6.3	336	1	YD48_METJA	Q97473	Methanococ
29	81	6.4	441	1	PSB1_CVATA	P43101	Cyanobacter
30	80.5	6.3	353	1	MT04_HELFP	G25443	Helicobacte
31	80.5	6.3	700	1	CAN2_MOUSE	G34929	Mus muscul
32	80.5	6.3	752	1	NECL_RAT	P28840	Rattus norv
33	80	6.3	953	1	YNM7_YEAST	P53927	Saccharomy

RESULT 1
GFP_AEQV1
ID: GFP_AEQV1 STANDARD: PRI: 248 AA.

AC P42212: Q17104:
DE 01-NOV-1995 (Rel: 32, Created)
DE 01-NOV-1995 (Rel: 32, Last sequence update)
DE 16-OCT-2001 (Rel: 40, Last annotation update)
DE Green fluorescent protein.

GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID:5100:
LN 1:
RN RP
RF SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE:92175527; PubMed:1347277;
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA Cormier M.J.:
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein.";
RL Gene 111:229-233(1992).

RP SEQUENCE FROM N.A.
RX MEDLINE:94185810; PubMed:6147953;
RA Iizuye S., Isuji F.:
RT "Aequorea green fluorescent protein. Expression of the gene and
RT fluorescence characteristics of the recombinant protein.";
RL FEBS Lett. 341:277-280(1994).

RP GFP-GENE-PROBE.
RX MEDLINE:93192221; PubMed:8448132;
RA Cody G.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.:
RT "Chemical structure of the hexapeptide chromophore of the Aequorea
RT green-fluorescent protein.";
RL Biochemistry 32:1212-1226(1993).

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE:96355665; PubMed:8703025;
RA Grime M., Cobitt A.B., Kallio K., Cross L.A., Tsien R.Y.,
RA Reibstahl S.J.:
RT "Crystal structure of the Aequorea victoria green fluorescent
RT protein.";
RL Science 273:1392-1395(1996).

RP X-RAY CRYSTALLOGRAPHY (1.0 ANGSTROMS).
RX MEDLINE:96294541; PubMed:9631087;
RA Yang F., Moss L.G., Phillips G.N., Jr.:
RT "The molecular structure of green fluorescent protein.";
RL Nat. Biotechnol. 14:1246-1251(1996).

RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
RX MEDLINE:98455509; PubMed:9782051;
RA Wachtel R.M., Eisliger M.A., Kallio K., Hanson G.T., Remington S.J.:
RT "Structural basis of spectral shifts in the yellow-emission variants
RT of green fluorescent protein.";
RL Structure 6:1267-1277(1998).

ALIGNMENTS

34	79.5	6.2	312	1	TRAX_GHLIR	Q84101	Chlamydia t
35	79.5	6.2	658	1	ADAS_CAVPO	P97275	Savia porce
36	79.5	6.2	860	1	SYL_ECOLI	P07815	Escherichia
37	78.5	6.2	263	1	PENK_MOUSE	P22005	Mus musculus
38	78.5	6.2	897	1	ITH3_RAT	Q63416	Rattus norv
39	78	6.1	451	1	PSRC_GHELU	Q06684	Chlamydom
40	78	6.1	684	1	EPGA_METHH	O27190	Metanaphaet
41	74	6.1	737	1	OPT1_DROME	P91679	Drosophila
42	73	6.1	2334	1	KAPA_BACSC	Q07833	Bacillus su
43	72.5	6.1	227	1	EXPL_ERWCA	P33882	Erwinia car
44	72.5	6.1	390	1	PYPC_SOTSO	Q50X05	Scafolobus
45	72.5	6.1	760	1	CAN2_RAT	Q07809	Rattus norv

[7]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
MOLIMES=96239303; PDBID:10220315;
Elisner M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.
"Structural and spectral response of green fluorescent protein
variants to changes in pH." (1999).
Biochemistry 38:5296-5301(1999).
- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE
BLUE CHEMILUMINESCENCE OF THE PROTEIN AECORIN INTO GREEN
FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO ON
RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PHOTOPROTEIN AECORIN.
ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
ABSORBANCE PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.
- SUBUNIT: MONOMER.
- TISSUE SPECIFICITY: PHOTOCYTES.
- P.M. CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED
OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
CYCLIZATION OF THE RESIDUES SER-DREYDROTYL-GLY.
- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
chimeric proteins of GFP linked to other proteins where it
functions as a fluorescent protein tag. GFP tolerates N-and C
terminal fusion to a broad variety of proteins. It has been
expressed in bacteria, yeast, slime mold, plants, protozoa,
zebrafish, and in mammalian cells. As a noninvasive fluorescent
marker in living cells, it allows for a wide range of applications
where it may function as a cell lineage tracer, reporter of gene
expression, or as a measure of protein-protein interactions.
- DATABASE: NAME-Protein Spotlight.
NOTE=Issue 11 of June 2001.
WWW="http://www.expasy.org/spotlight/articles/spot101.html".
This SWISS-PROT entry is copyright. It is produced under a license
between the Swiss Institute of Bioinformatics and the EMBL outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in its
modified and this statement is not removed. Usage by for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/usage/>)
or send an email to license@sib-sb.ch.
EMBL: M62654; AAA27722.1;
EMBL: M62654; AAA27722.1;
FMBL: L29345; AAA58246.1;
PDB: 1GFL; 1JAN-97
PDB: 1EWA; 08-NOV-96
PDB: 1EMB; 16-JUN-97
PDB: 1EMC; 20-AUG-97
PDB: 2EMD; 20-AUG-97
PDB: 1EME; 20-AUG-97
PDB: 1EMF; 20-AUG-97
PDB: 1EMG; 12-MAY-99
PDB: 1EMK; 20-AUG-97
PDB: 1EWL; 20-AUG-97
PDB: 1EMM; 20-AUG-97
PDB: 2EMN; 20-AUG-97
PDB: 2EMO; 20-AUG-97
PDB: 1BFP; 07-JUL-97
PDB: 1VFP; 28-OCT-98
InterPro: IPR000756; Green_fl_protein.
Pfam: PF01553; GFP_1
PRINTS: PK01229; GFP+JORESCENT.
ProDom: P0013756; Green_fl_protein: 1.
Luminescence: 3D-structure.
SITE 65 67
FT VARIANT 100 100 F > Y.
FT VARIANT 108 108 T > S.
FT VARIANT 141 141 L > M.
FT VARIANT 219 219 V > L.
FT VARIANT 25 25 R > G (IN REF. 2).
FT CONFLICT 157 157 Q > P (IN REF. 2).
FT CONFLICT 172 172 E > K (IN REF. 2).
FT CONFLICT 172 172 E > K (IN REF. 2).
FT SEQUENCE 238 AA; 26886 MW; EA5A5E21F8FE6E05 CDS64.

```

Query Match          98.74; Score 1254; DB 1; Length 238;
Rest Local Similarity 98.34; Pred. No. 6,3e-100;
Matches 234; Conservative 1; Mismatches 3; Gaps 0; Caps 0;

QY      1 MSKGEELTGVP-LVFLGDGVNCHRFVSQEGDAIGYGLTKLP-CTTICKLHVPRPZII 60
DB      1 MSKGEEUFTGWFILVELDGVNHGRFVSQGGEGDGYGLDLNKFLCTGKLHPVAPTL 60
QY      51 VTTLSYGVOCESRYPPDHMKHDPFKSAMPEGYVOERTFFPKDDGNKYTRAEVKFSPTLV 120
DB      51 VTTLSYGVOCESRYPPDHMKQHDFKSAMPEGYVOERTIFFPKDDGNKYTRAEVKFSPTLV 120
QY      121 NRIELGIDKEGDNILGHLEYNASHNVNYTMADCKNGIKVNEKISHNIETDSGYOIAI 180
DB      121 NRIELGIDKEGDNILGHLELTNYNSHNVTMADCKNGIKVNEKISHNIETDSGYOIAI 180
QY      131 HYQONTPIGSPVLLPNHHYLSTQSALSQDKPKEDHMVLGFVTAAGITHMGDELYK 238
DB      131 HYQONTPIGSPVLLPNHHYLSTQSALSQDKPKEDHMVLGFVTAAGITHMGDELYK 238

RESULT 2
SYL_XYLFA STANBARK; ERT: 879 AA.
AC Q9PB68.
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 43, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (leucine--tRNA ligase) (leuRS)
GN LEUS OR XZ1215.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
GX Xylella
GX NCBI_TaxID=2371;
RP [ ]
SEQUENCE FROM N.A.
PC STRAIN=545C;
MEDLINE=20355717; PubMed=16910347;
RA Simpsons A.C.G., Weinbach P.C., Arruda P., Abreu F.A., Agencio M.,
RA Alvarado R., Alves M.C., Araya J.E., Bal G.S., Baptista C.S.,
RA Barros M.B., Bonaccorsi E.P., Bordin S., Bove J.M., Beloncs M.F.S.,
RA Castro M.R.P., Camargo A.A., Camargo L.F.A., Carraro D.N., Carreira H.,
RA Colatto K.B., Colombo C., Costa F.P., Costa M.C.K., Costa Neto G.A.,
RA Coutinho I.L., Cristofani X., Dias-Neto E., Decena C.G., El-Borly H.,
RA Fadlouni A.P., Ferreira A.J.S., Ferreira V.O.A., Ferriz J.A.,
RA Praga J.S., Franco S.G., Franco M.C., Franco M., Furian L.R.,
RA Carrier M., Goldman G.H., Goldman M.S., Gomes S.A., Gruber A.,
RA Ho P.S., Hochreisel T.D., Jurgenova M., Keuter F.L., Kitajima J.P.,
RA Klueger J.B., Kruman E.E., Laigret F., Lambais M.K., Leite L.C.,
RA Lemos E.G.N., Lenks M.V.F., Lopes S.A., Lopes C.M., Machado J.A.,
RA Machado M.A., Madeira A.M.N., Madeira E.M.F., Marino C.L.,
RA Marques M.V., Martins A.M.L., Martins E.M.F., Marukuma A.Y.,
RA Menck C.F.N., Miracco E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.C., Netto L.E.S.,
RA Nhapi A., Jr., Nohrda F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto S.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero D.A.,
RA Guadalupe R.B., Roberto F.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshakmo M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vertore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RL "The genome sequence of the plant pathogen xylella fastidiosa."
RL Nature 406(15)-159(2006).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) -> AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collection

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/ebis/sequence/> or send an email to license@ebi.ac.uk).

EMBL: AE004031: ANP4975.1: AL1.1.KT.
InterPro: IPR023024: euk-ERASYNTH.
InterPro: IPR023025: ERNA-synTH.
InterPro: IPR023026: ERNA-synTH.
Print: PF00133: ERNA-synTH.
Print: PF00245: ERASYNTH.
TIGRFS: TIGR00394: ERASYNTH.
PROSITE: PS0178: AA-ERNA-ERASE.1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ATP-binding.
KW Complete proteome.
FT SITE 45 55 "HIGH" REGION.
FT SITE 537 64 "KSKS" REGION.
FT BINDING 640 640 ATP (BY SIMILARITY).
SO SEQUENCE 879 AA: 9976 MW: 96264624.24cHE Cys44:

Query Match: 73% Score 65.01 ID: E0618.8728
Best Local Similarity: 23.24; Pred No. 0.76
Matches: 46; Conservative: 28; Mismatches: 17; Gaps: 10;

QY 49 TTGKLPVMPPLVTTISYGVGCFSTYTHMKRHFESAKFBNVGFETTFKQDARY--106
DB 321 TNEGLPVWANEVLMAGTGAANVAPCHGQRQEF--AKKYGIDTFQVIAKEPKNGDE 377
QY 107 -----KIRAEVKEEDTIVNRIEKGIDPEKDNLSHKLENYNSNRYI 152
DB 378 STWEPDWRWYADKIR--EFE--GNSAEVGLDYDAFVLAHPE----- 421
QY 153 MACKQKNC-IVNFKIRN-ENGQVQACVYQNTPI-----NUGVQLPEN 196
DB 422 --KQSGQGRVYRLK----DWVSQRKQKRIPLVGYDTGAVPVPEGLVLPEN 471
QY 199 -HYLSTOSALKLPINEX 215
DB 475 VAFSGTGP-KTQPEWRK 452

RESILI 3
NIFD_CLOPA
13 NIFD_CLOPA STANDARD: PRT: 543 AA.
AC P00467;
DT 21-JUL-1986 (Ref. 01, Created)
DI 01-APR-1988 (Ref. 07, Last sequence update)
DI 15-JUL-1999 (Ref. 38, Last annotation update)
DE Nitrogenase molybdenum-iron protein alpha chain (EC 1.18.6.1)
DE (Nitrogenase molybdenum-iron protein alpha chain)
GN NIFD.
CS Clostridium pasteurianum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
CX NCBI_TaxID=1561;
RW [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=M7231195; PubMed=1473447;
RX Wang S.-Z., Chen J.-S., Johnson J.L.;
RX "Nucleotide and deduced amino acid sequences of nifD, encoding the
alpha-subunit of nitrogenase molybdenum-iron protein of
Clostridium pasteurianum";
RL Nucleic Acids Res. 15:4935-4945(1987).
RN [2]
RP SEQUENCE OF 1-179.
RX MEDLINE=R2030659; PubMed=226551;
RX Hase T., Nakao T., Matsubara H., Sumit W.;
RX "Correspondence of the larger subunit of the K. ferroplasma to
the clostridial nitrogenase to the nif D gene products of other N2 fixing
organisms";
RL J. Biochem. 90:295-298(1981).
RN [3]

RP SEQUENCE OF 1-166 FROM N.A.
RX MEDLINE=R6168010; PubMed=3457003;
RX Chen K.-K., Chen J.-S., Johnson J.L.;
RX "Structural features of multiple nif-like sequences and very biased
codon usage in nitrogenase genes of Clostridium pasteurianum";
RI J. Bacteriol. 156:1162-1172(1996).
RN [4]
RP SEQUENCE OF 525-533 FROM N.A.
RX MEDLINE=R8294019; PubMed=2840948;
RX Wang S.-Z., Chen J.-S., Johnson J.L.;
RX "Distinct structural features of the alpha and beta subunits of
nitrogenase molybdenum-iron protein of Clostridium pasteurianum: an
analysis of amino acid sequences";
RL Biochemistry 27:2850-2810(1988).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=R3344379; PubMed=8933705;
RX Kim J., Woo D., Rees D.C.;
RX "X-ray crystal structure of the nitrogenase molybdenum-iron protein
from Clostridium pasteurianum at 3.0-A resolution";
RL Biochemistry 32:7104-7115(1993).
CC -!- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE
CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.
CC -!- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP
+ 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS THAT BINDS
30-32 FE, 2 MO, AND INORGANIC SULFUR.
CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: M21547: AAA83531.1;
EMBL: Y00155: AAA68349.1;
EMBL: M23380: AAA21265.1;
FIR: S07389: NIFK.NMA.
FIR: M10: 31-OCT-93.
InterPro: IPR00318: Nitrogenase compl.
InterPro: IPR00510: Oxidized nitrogenase.
Ffam: PF0348: oxidized nitro.
TIGRFS: TIGR01282: nifH.1
PROSITE: PS00390: NITROGENASE_1_2; 1;
PROSITE: PS00699: NITROGENASE_1_1; 1;
KW Oxidoreductase; Nitrogen fixation; Molybdenum; Iron-sulfur;
KW 3D-structure.
FT INIT-MET 0 0
FT CONFLICT 41 41 P -> K (IN REF. 2).
FT CONFLICT 94 94 N -> D (IN REF. 2).
FT HELIX 3 9
FT TURN 10 11
FT HELIX 14 21
FT TURN 22 23
FT STRAND 24 27
FT TURN 44 45
FT HELIX 53 62
FT TURN 63 64
FT STRAND 65 65
FT TURN 66 67
FT STRAND 68 73
FT HELIX 72 76
FT TURN 81 82
FT STRAND 83 88
FT HELIX 93 95
FT TURN 101 102
FT STRAND 105 106
FT HELIX 111 115
FT TURN 116 117

CC AND FETAL TISSUES, XENIN IS FOUND IN GASTRO, ENDOCRINE AND CENTRAL
CC MUCOSA. CIRCULATES IN THE BLOOD. SEEMS TO BE CONFINED TO SECRETORY
CC ENDOCRINE CELLS.
CC -1- DEVELOPMENTAL STAGE: XENIN IS RELEASED INTO THE CIRCULATION AFTER
CC A MEAL.
CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (TREP-ASP DEVIANS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL institution -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch)
CC
CC EMBL: U24105; AAB70879.1; 1;
CC Genes: HGNC:2230; CABA.
CC XIM; 601924; 1;
CC InterPro: IPR001460; WD40
CC Pfam: PF00409; WD40; 7;
CC PRINTS: PR00130; GPROTEINRPT.
CC ProDom: PD00018; WD40; 6;
CC SMART: SM00320; WD40; 6;
CC PROSITE: PS00676; WD-REPEATS_1; 1;
CC PROSITE: PS00682; WD-REPEATS_2; 6;
CC PROSITE: PS00294; WD-REPEATS_REGION; 1;
CC Transprot: Protein transprot: Golgi stack; Endoplasmic reticulum;
KW Membrane: Repeat: WD repeat; Hormone;
FT PEPTIDE 1 25 XENIN.
FT PEPTIDE 1 35 PRORENIN.
FT REPEAT 1 47 WD 1.
FT REPEAT 2 49 79 WD 2.
FT REPEAT 3 93 121 WD 3.
FT REPEAT 4 133 163 WD 4.
FT REPEAT 5 203 233 WD 5.
FT REPEAT 6 247 277 WD 6.
FT REPEAT 7 289 319 WD 7.
SQ SEQUENCE 1224 AA; 138331 MW; 5A3B08845D8F15D CRC64.

Query Match 7.1%; Score 90; DB 1; Length 1224;
Best Local Similarity 25.9%; Pred. No. 7.3;
Matches 41; Conservative 26; Mismatches 59; Indels 32; Gaps 7;

QY 90 EGVQRTFFKDDGNYKTRAEVKEFSDIYNKIF-LKIDFKEDNLTGKIKFYNKSH 146
DB 632 KGYFEVALHFVQDE---KTFSLALGCGNLTALAAKALGCKN*WFKLGEVALQGNHQ 586
QY 149 NVYIMADKQKNGIKNF-----KIRHNDSDGVO---LADHYQNTPIQLPVLFPD 197
DE 689 IVEMQRTKTKDKVSLFLITCNLEKIRMKKIATIRKMSGRYONALYAG----- 241
QY 198 NHYISTQSALSQDPNEKDHVLLQFVTAAGITHGND 235
DB 742 ---VSEVRILKCGQKS---LAVLTAA--THGLIDE 749

RESULT 5
IC SYL_RASIN STANDARD; P87; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Eukaryotes)
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_TaxID:727;
PP (1)
RN
PC STRAIN:RD / KW20 / ATCC 51907;
RX MEDLINE:95350630; PubMed:7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kervavage A.P., Ruit C.J., Tomb J.E., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu D.T., Glodt A., Keiley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.E.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saidek D.M., Brandon R.C.,
RA Fine L.D., Pritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Guelm C.D., McDonald L.A., Small K.V., Fraser C.M., Smith H.H.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RA R1";
RA Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-T AMINOCYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL institution -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch)
CC
CC EMBL: J05774; AAC23581.1;
CC XIGR; H10921;
CC InterPro: IPR02302; Leu-tRNA-synthet.
CC InterPro: IPR02303; tRNA-synth_1a.
CC InterPro: IPR01412; tRNA-synth_1.
CC Pfam: PF00133; tRNA-synth_1; 1;
CC PRINTS: PR00985; TRNASYNTHLE.
CC TIGRFAMS: T1GR00396; leus_hact; 1;
CC PROSITE: PS00178; AA-tRNA-LIGASE_I; 1;
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 519 523 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA; 97750 MW; E593304F5BAC6F7 CRC64;

Query Match 7.3%; Score 89.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 5.2;
Matches 45; Conservative 26; Mismatches 57; Indels 47; Gaps 9;

QY 49 TTGKLPVWPETLTLSYGQCFRYPDHMKRDFKSAPEGYVOERTIFKDDGNYKT 108
DB 314 TGUKLPI-WVANFVLMYHGVAVNPAHDCRD--PEFAKYSLPKQVLAFLADEEID 370
QY 109 RAEVKFEQDLYNRLE-KGIDFKEDQNLGHKLEYNSHNHYIMADK-QRNS-----KVN 164
DB 371 TKQAFVHGKLVNSDEDFGKNF--DGAENG-----IADK-EKLVAGKQVN 414
QY 165 FKIEH-----NIENGQVLADHYQNTPIGDPVLLFNHLS-STQSAL 207
DB 415 YKRDWGVSKRYWGAFLPMCTLENGDVPA-----PMEDLPIIDPDVWGVKRSPI 467
QY 206 SKDPN 212
DB 465 NADPN 472

RESULT 6
IC HIS7_BACAI STANDARD; P81; 353 AA.
AC P57203;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histidine biosynthesis bifunctional protein HisB (includes:
DE Histidinol-phosphatase (EC 3.1.3.15); Imidazoleglycerol-phosphate
DE dehydratase (EC 4.2.1.19) (IGPD)).
GN HISB OR B0102.

OS *Bacillus aphidivorus* (subsp. *Aphythiopsis aphidivorus*) (Aphythiopsis aphidivorus)
OS symbiotic bacterium.
OC Bacteria: Proteobacteria: gamma subdivision: Bacteri.
CX NCBI_taxid=118099;
RN [1]
RN SEQUENCE FROM N.A.
PC STRAIN=Tokyo 1998;
PC MEDLINE=20445173; PubMed=10994077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.
RI "Genome sequence of the endocellular bacterial symbiont of aphids
Baccharis sp. AFS."
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: D-erythro-1-(imidazol-4-yl)phosphono-3-
phosphate -> 3-(imidazol-4-yl)-2-oxopropyl phosphate + H(2O) - 1-histidinol
+ phosphate.
CC -!- PATHWAY: Histidine biosynthesis: eighth step.
CC -!- PATHWAY: Histidine biosynthesis: eighth step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE HISTIDINOL-
PHOSPHATE FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/information/>
or send an email to license@sib-sib.ch).
CC EMBL: AF061115; BAR12821.1;
DR InterPro: IPR000807; IPR01;
DR Pfam: PF06475; IPR01.1;
DR ProDom: PD002282; IPR01.1;
DR TIGRfam: TIGR01261; EMBL_NORM: 1;
DR PROSITE: PS00354; IPR_DEHYDRATASE_1; 1;
DR PROSITE: PS00355; IPR_DEHYDRATASE_2; 1;
KW Histidine biosynthesis; Multifunctional enzyme; Lyase; Hydrolase;
KW Complete proteome.
FT DOMAIN 1 164 HISTIDINOL-DEHYDRATASE
FT DOMAIN 165 453 IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE
SQ SEQUENCE 453 AA: 40845 MW: 5048275.600
Query Match 64% Score 84.00 E: 1e-141 Length 453
Best Local Similarity 44.0% Pred No. 2.0
Matches 49; Conservative 44; Mismatches 44; Gaps 10;
CY 63 LSTGVQESRYDQPKKHIDPKKSNMPSYVQVE-----FKSNMPSYVQVE---RTT 98
DB 62 TSEFFLOPS- TALEFMISYERS---EQTVDGLLGHIFDQPTGPKKIRHIFNL 111
CY 115 -----EQT- CARRHKKIDKEGNSLHMKYNSNHYIMAKQKN 153
DB 116 EMILKKSVVIGKRODMLSNLKKIKKKEG-CKNKHITVITAKHYATIDPIK 174
CY 160 GKVNFKPIHNI-----DGGVQLADHYQNTGIDKPVLPKHLYST-SALSKTENEK 214
DB 175 ETKSVIKVWLDSELSKIDGKVFTHMLGVSVKSGTGM-----NISVQGLDGLT 229
CY 215 -FHVWVLLG--FVTAAGTIGHM 233
DB 230 IEDTGIVLGEALLQALGKKNL 251

RESULT 7

D152_HAEIN
ID D152_HAEIN STANDARD: PRI: 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN H10917.
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria: gamma subdivision: Pasteurellales.
CX NCBI_taxid=727;
RN [1]
RN SEQUENCE FROM N.A.
PC STRAIN=Re / KW20 / ATCC 51907;
PC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Klevorick A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott C.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ufferback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Ouellet C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RI "Whole-genome random sequencing and assembly of Haemophilus influenzae
Re."
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/information/>
or send an email to license@sib-sib.ch).
CC EMBL: U12773; AAC22575.1;
DR EMBL: H10917.1;
DR InterPro: IPR005184; Bac_surfac_D15.
DR Pfam: PF01103; Bac_Surface_Ag; 1;
DR Actinon: Outer membrane: Signal: Complete proteome.
FT SIGNAL 1 19 POTENTIAL
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15
SQ SEQUENCE 795 AA: 87478 MW: 885691FC22EED44 CRO64;
Query Match 64% Score 87.50 E: 1e-141 Length 795
Best Local Similarity 22.4% Pred No. 7.1
Matches 49; Conservative 44; Mismatches 73; Indels 63; Gaps 12;
CY 64 LSTGVQESRYDQPKKHIDPKKSNMPSYVQVE-----FKSNMPSYVQVE---RTT 98
DB 427 IGVCTESGISVQASVQKQNFQSGAASVAGTKNDYGTSVNAGYTPYTKGVSIQNV 486
CY 99 FPKDGNVKTAEVKPEKGGUTANNRICKKIDKEGNSLHMKYNSNHYIMAKQKN 147
DB 99 FPKDGNVKTAEVKPEKGGUTANNRICKKIDKEGNSLHMKYNSNHYIMAKQKN 147
CY 487 FPKVNSKSESSSRKTKLYGKNVTLGFPVNNNSYVGLCHTYNKISREALEYN--- 512
DB 487 HRYVIMADKQK-NGTKVN-----FKIRHN-----IEDGVQADHYQNTGIDKPVLPKH 195
CY 547 RNLYIQSMKFKGNIKTDFSEFGNNVNSLNGVPEPKGVKAS-----LS-GVVTI 593
DB 547 RNLYIQSMKFKGNIKTDFSEFGNNVNSLNGVPEPKGVKAS-----LS-GVVTI 593
CY 596 P--DNHLYLTQSALSKIDPKKIDHMLVLLSPVTAAGTIGHM 632
DB 594 PGSDNRYKXLSADVQGFYDIDKCHLWWSAKASAGVANG 632
RESULT 8
D152_HAEIN
ID D152_HAEIN STANDARD: PRI: 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).

OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus
CA NCBI_taxonomy:727;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN: Serotype B;
EX MEDLINE:95255676; PubMed:7737523;
RA Fick P.S., Loomore S., Chow P., Thomas W.B.,
RT The sequencing of the 43-kDa D15 protective surface antigen of
R Haemophilus influenzae.";
R1 Gene 156:97-99(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN: Mima and Egge / Serotype B;
RX MEDLINE:97427952; PubMed:9284145;
RA Loomore S.M., Yang Y.P., Coleman D.C., Shortland J.M., England D.H.,
RT *Outer membrane protein D-5 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease".
R1 Infect Immun 65:3701-3707(1997).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN: BRINGS TO THE SURFACE ANTIGEN D15 FAMILY
CC (1) - SIMILARITY: BRINGS TO THE SURFACE ANTIGEN D15 FAMILY
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U13961; AA085045.1;
DR EMBL: U60832; AA061974.1;
DR EMBL: U60833; AA061976.1;
DR InterPro: IPR000184; Rat_sur:A4_015.
DR Pfam: PF01103; Bac_surface.Aq: 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA: 87675 MW: 2793DE53869AF1E CPC64;

Query Match 5.9%; Score 87.5; DB 1; Length 797;
Best Local Similarity 22.4%; Prod. No. 7.1;
Matches 45; Conservative 34; Mismatches 7; Indels 43; Gaps 14;

QY 64 ISYGVQCFSPDKKHEF-----PKSAMPEGVQCF-----KLYFVS 147
DY 427 IYGTESGISTYQASVKQNFELTGTGAASVAGTKNDYGVSVAGVATYFTKGVSYGTRY 426
QY 94 PFKDGNYKTRAEVAFEGTILVNRTELKSTDEKENGZ--KSH-----KLYFVS 147
DY 447 FTENVNKSQTSNYSKYKTTGNSVLTGPPVNNKSNVYVAGVATYFTKSNFALEYS- 443
QY 148 HNVYTMASQK-NGIKV-----FKRN-----FQGVJAHVQNTGDFPVLL 146
DY 543 RNLVQSKEKFGNG-KTNDPFGWNGNSLVAFYFTKSNFALEYS- 540
QY 196 P--QNEY--STQASQKQNEKEHMYLGGVTAAGTILS 214
DY 634 PQSKNYKLSADVQCFPTLDFDRLHVVYSKASQAYANG 632

RESULT 5
COPA_BOVIN STAN:ABD. PRT: 1224 AA.
AC Q27954;
DI 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coatomer alpha subunit (Alpha-coat protein) (Alpha-coat protein)
DE (HEP-COP) [Contains: Xenin (Xenopsin-related peptide); Proximin...]

GN COPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovine; Bos.
CA NCBI_taxonomy:9913;
RN (1)
RP SEQUENCE FROM N.A.
RC Wicland F.1;
CC FUNCTION: THE COATOMER IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS
CC TO DILYSINE MOTIFS AND REVERSLY ASSOCIATES WITH GOLGI VES-
CC CLES, WHICH FURTHER MEDIATE BIOSYNTHETIC
CC PROTEIN TRANSPORT FROM THE ER VIA THE GOLGI UP TO THE TRANS GOLGI
CC NETWORK. COATOMER COMPLEX IS REQUIRED FOR BUDDING FROM GOLGI
CC MEMBRANES, AND IS ESSENTIAL FOR THE RETROGRADE GOLGI-TO-ER
CC TRANSPORT OF DILYSINE-TAGGED PROTEINS. IN MAMMALS, THE COATOMER
CC CAN ONLY BE RECRUITED BY MEMBRANES ASSOCIATED TO ADP-RIBOSYLATION
CC FACTORS (ARF5), WHICH ARE SMALL GTP-BINDING PROTEINS. THE COMPLEX
CC ALSO INFLUENCES THE GOLGI STRUCTURAL INTEGRITY, AS WELL AS THE
CC PROCESSING ACTIVITY, AND ENDOCYTIC RECYCLING OF LDL RECEPTORS (BY
CC SIMILARITY).
CC (1) - FUNCTION: XENIN STIMULATES EXOCRINE PANCREATIC SECRETION. IT
CC INHIBITS PANCREATIC-STIMULATED SECRETION OF ACID, TO INDUCE
CC EXOCRINE PANCREATIC SECRETION AND TO AFFECT SMALL AND LARGE
CC INTESTINAL MOTILITY. IN THE GUI, XENIN INTERACTS WITH THE
CC NEUROTENSIN RECEPTOR (BY SIMILARITY).
CC (1) - SUBUNIT: OLIGOMERIC COMPLEX THAT CONSISTS OF AT LEAST THE ALPHA,
CC BETA, BETA', GAMMA, DELTA, EPSILON AND ZETA SUBUNITS.
CC (1) - SURCELLULAR LOCATION: THE COATOMER IS CYTOPLASMIC OR POLYMERIZED
CC ON THE CYTOPLASMIC SIDE OF THE GOLGI, AS WELL AS ON THE
CC VESICLES/BUDS ORIGINATING FROM IT (BY SIMILARITY). XENIN IS
CC SECRETED (BY SIMILARITY).
CC (1) - SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X96768; CAA65543.1;
DR InterPro: IPR001680; WD40
DR Pfam: PF00400; WD40: 7.
DR PRINTS: PR00320; GPROTEINRNT.
DR PROSITE: PS000018; WD40: 5.
DR SMART: SM00320; WD40: 6.
DR PROSITE: PS00678; WD_REPEATS_1; 1.
DR PROSITE: PS00082; WD_REPEATS_2; 6.
DR PROSITE: PS00294; WD_REPEATS_REGION; 1.
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum;
KW Membrane; Repeat; WD repeat; Hormone.
FT PESTIDE 1 35 XENIN.
FT PESTIDE 1 35 PROTEIN.
FT REPEAT 7 37 WD 1.
FT REPEAT 49 79 WD 2.
FT REPEAT 51 123 WD 3.
FT REPEAT 133 163 WD 4.
FT REPEAT 203 233 WD 5.
FT REPEAT 247 277 WD 6.
SQ SEQUENCE 1224 AA: 138456 MW: C9350BF2AC00693D CPC64;

Query Match 6.9%; Score 87.5; DB 1; Length 1224;
Best Local Similarity 22.4%; Prod. No. 12;
Matches 54; Conservative 33; Mismatches 65; Indels 69; Gaps 12;

QY 44 LKFCITGKLPV-----PWTLVTL-SYGVQCFSP-----Y 74
DY 549 IKAVITGSGYGLTLDLPYIVTRKGNVNVCLDRECRPVLTIDPTEPKFLALNNRY 608
QY 75 PD--HKRRH-----DFFKSAMPEGVQCFPTLDFDRLHVVYSKASQAYANG 125

```

CC 609 DEVHVRNRKAKGKSLIAFAKSKSTFEVALRPTDSE - KTFSEALAEENRINIA LAA 947
CC 126 KGIDFKEDGNGILGKIKLENNENNVIMARKKNITRYNU-- --KTFPENETENNV 170
CC 666 KALDRKNCWKEGCEVALD_GNRGIVEMCVLRKIKLEKSPFLVLTN_GKIKAKKOKIAEL 245
CC 177 Q--LAHRYQONTGCGSVLLPNNHVI STCSALSKLPKPKRKHRRVLAQFVLAQGLHRRD 284
CC 726 EKMSGHYUNALYGD - - - - -VSRKVKIKRNGQKRS-- -LAVLAA--THGLD 266
CC 235 E 235
CC 269 E 269

RESULT 10
PROE_YEAST
AC P21951:
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2002 (Rel. 45, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
DE polymerase II subunit A).
GN POL2 OR DUN2 OR YNL262W OR NUG25.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NBLTaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.A. AND SEQUENCE OF 1214-1221.
RX MEDLINE 96381771; PubMed=5740425;
RA Sen-Gupta M., Lyck R., Selig U., Naedertal K.R., Hopmann F.H.,
RA Morrison A., Ataki H., Clark A.R., Hamatake R.K., Saito A.,
PT "A third essential DNA polymerase in S. cerevisiae".
PL Cell 62:1143-1151(1990);
RN (2)
RP SEQUENCE OF 1-2-21 FROM N.A.
PC STRAIN=S288c / FY1679;
RX MEDLINE=96310631; PubMed=5740425;
RA Sen-Gupta M., Lyck R., Selig U., Naedertal K.R., Hopmann F.H.,
RA Morrison A., Ataki H., Clark A.R., Hamatake R.K., Saito A.,
PT "The sequence of a 24,152 bp segment from the left arm of chromosome
PT XIV from Saccharomyces cerevisiae between the PHU1 and the POL2
RE genes".
RL Yeast 12:505-514 (1995).
RN (3)
RP TEMPERATURE SENSITIVE MUTANTS.
RX MEDLINE=92144663; PubMed=1537345;
RA Araki H., Ropp P.A., Johnson A.L., Taneston L.E., Morrison A.,
RA Saito A.,
PT "DNA polymerase II, the probable homolog of mammalian DNA polymerase
PT epsilon, replicates chromosomal DNA in the yeast Saccharomyces
PT cerevisiae".
RL FMO 3: 11733-740(1992).
CC -1- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
CC REPLICATION.
CC -1- CATALYTIC ACTIVITY: N-deoxyribonucleoside triphosphate - N-deoxyribo-
CC + (DNA)(N).
CC -1- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 130 kDa, 34 kDa, 40
CC kDa, AND 29 kDa).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNIT B AND C.
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. HIGH
CC SIMILARITY WITH MAMMALIAN DNA POLYMERASE EPSILON.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use of non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.eth.ch/annoucy/
CC or send an email to license@isb.eth.ch).
CC
CC EMBL: M40456; AAA68711; 1;
CC EMBL: X92494; CAA63235; 1;
CC EMBL: Z71535; CAA61691; 1;
CC FIK: A36028; AF6028;
CC SGD: SC005206; PO22;
CC InterPro: IPR002064; DNA_pol_B;
CC Pfam: PF01136; DNA_pol_B; 1;
CC PRINTS: PR00106; DNAPOLB;
CC SMART: SM00486; POLB; 1;
CC PROSITE: PS00116; DNA_POLYMERASE_LB; FALSE_NEG.
KW Transferrase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Zinc-finger; Nuclear protein;
PT 2NLFNG 2108 2151 POTENTIAL.
FT VARIANT 644 644 M -> I (IN POL2-5 IS MUTANT).
FT VARIANT 710 710 P -> S (IN POL2-18 IS MUTANT).
SQ SEQUENCE 2222 AA; 255669 MW; CECDE2AB147D65B CRC64;
Query Match. 6.9%; Score 67.5; DB 1; Length 2222;
Best Local Similarity 28.2%; Pred. No. 24;
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;
QY 53 LPVWMP-TLVTTISYGVQSPSKYPOHM-----KRHDFKSAMPGYV---QERTI 98
DB 883 LPKSFETVFFLENGKKLYLSYPSKLNRYVHQFTNHCYOSLKDPNLYIETHSEMTI 942
QY 99 FFKDGGNYKTR--AYKPGSDTLVNR-----IELKGIDFKEDNTGKLEKLYNN 146
DB 943 FFEVDGPKYK--LPSSKEEGKIKKRYAVENEGSLALGKELKRGKFL---QLIKNQ 999
QY 147 S--HNVLMD 155
DB 1500 SDIFKVFLEGE 1010
REMARK 11
TRNA_CAME
CC TRNA_CAME STANDARD; PRT; 357 AA.
AC QEP242:
DT 15-JUN-2002 (Rel. 41, Created);
DT 15-JUN-2002 (Rel. 41, Last sequence update);
DT 15-JUN-2002 (Rel. 41, Last annotation update);
DE TRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (CPNA(M-5-05))-
DE methyltransferase (MONT).
GN TRNA OF CJC810.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NBLTaxID=197;
RN (1)
RP SEQUENCE FROM K.A.
RX MEDLINE=92150912; PubMed=1048204;
RA Parkhill J., Wren B.W., Mangai K., Ketley J.M., Churcher C.,
RA Bastein D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jacobs K., Karyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Randleam M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall R.G.,
PT "The genome sequence of the food-borne pathogen Campylobacter jejuni
PT reveals hypervariable sequences".
RL Nature 431:665-668(2000).
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
CC 54 (M-5-U54) in all tRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing thymine.
CC -1- SIMILARITY: BELONGS TO THE RNA M5C METHYLTRANSFERASE FAMILY. TRNA
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```


Query Match: 6.6% Score 94.5; DP 11; Length 799;
Best Local Similarity 21.0%; Pred. No. 11;
Matches 47; Conservative 38; Mismatches 62; Indels 57; Gaps 9;

QY 4 GELATIN--GVVPLVGLGVGNGH-----EVSYSRFGKQVATKTKTKTKCTGGKLVLP 56
DB 423 GEMHITGVGVVTEVTELSGQTHHSKSNFLNRAPEKDTF-INREVLSEFKLNG 440
QY 57 WPLVLT-----SYSGVCFES-EYDHRKRRHFKSAGEYVQVET-FFKDCGNYKT 139
DB 481 EYLVPSFEPNKGDCGRCVSEKKAQYDAVDEKLANRETLSEDDV-----DQVPSL 53
QY 109 RAEV-----AFPGDTLVNRELKQIDKFNFA----- 134
DB 538 FALGAGELAEISAFETQTLFRLVLAQKQIKSGHSTETKQKVSQHSQSGKQKLEK 507
QY 135 NLGHRKLEYNSHNYVMKUKKNG-KVNEKIKENLEDAVQ----- 174
DB 556 YELWTRID-KY-KYTRGVSTVSTMSYKKAQDAVPSF 143
QY 556 YELWTRID-KY-KYTRGVSTVSTMSYKKAQDAVPSF 143

RESULT 15

ITB3_MESAU STANDARD; 3 FT; 896 AA.

AC P97280: 15-JUL-1998 (Ref. 36, first sequence update)
DT 15-JUL-1998 (Ref. 36, last sequence update)
FT 16-OCT-2001 (Ref. 49, last annotation update)
DE Inter-alpha-trypsin inhibitor heavy chain B1 precursor (ITI heavy
DB Chain B3) (6C3).
GN Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;
OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Cricetinae;
OC Mesocricetus
OX NBL_TaxID:10936;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=97420668; PubMed=9275673;
RX Nakatani T., Suzuki Y., Yamamoto T., Sinohara H.;
RT "Molecular cloning and sequencing of cDNAs encoding three heavy-chain
RT precursors of the inter-alpha-trypsin inhibitor in Syrian hamster;
RT implications for the evolution of the inter-alpha-trypsin inhibitor
PT heavy chain family."
PT J. Biochem. 122:71-82;1997).
RN [2]
RP SEQUENCE OF 31-59; 446-472 AND 504-523, AND SUBUNIT 1.
RC TISSUE=Plasma;
RA MEDLINE=97016241; PubMed=8664957;
RX Yamamoto T., Yamamoto K., Sinohara H.;
RT "Inter-alpha-trypsin inhibitor and its related proteins in Syrian
RT hamster urine and plasma."
RX J. Biochem. 120:145-152;1996).
CC 1-1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONIN IN SERUM OR AS A
CC BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX FACTORS,
CC INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE
CC LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE
CC ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY
CC SIMILARITY).
CC 1-1- SUBUNIT: 1-ALPHA-1 PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM
CC ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN,
CC BIKUNIN. INTER-ALPHA-INHIBITOR (I-ALPHA-I) IS COMPOSED OF H1, H2
CC AND BIKUNIN. INTER-ALPHA-LIKE INHIBITOR (I-ALPHA-LI) OF H2 AND
CC BIKUNIN. AND PRE-ALPHA-INHIBITOR (P-ALPHA-I) OF H3 AND BIKUNIN.
CC 1-1- 27M: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CHONDROITIN
CC 4-SULFATE BRIDGE TO THEIR C-TERMINAL ASPARTATE (S)
CC SIMILARITY).
CC 1-1- SIMILARITY: BELONGS TO THE ITB3 FAMILY.
CC 1-1- SIMILARITY: CONTAINS 1 VWA DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between: the Swiss Institute of Bioinformatics and the EMBL Outstation.

CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC -----
CC EMBL: C89287; HAAT3940.1; -
CC InterPro: IPR02035; VWEA.
CC Pfam: PF00692; VWA: 1.
CC SMART: SM00327; VWA: 1.
CC PROSITE: PS00234; VWA: 1.
KW Serine protease inhibitor; Repeat; Signal; Multigene family;
KW Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 30 BY SIMILARITY.
FT CHAIN 31 646 INTER-ALPHA-1-TRYPsin INHIBITOR HEAVY CHAIN
FT FT H3. SIMILARITY.
FT PROPEP 647 886 BY SIMILARITY.
FT DOMAIN 279 439 VWA.
FT CARBOHYD 88 89 N-LINKED (GLNAC...) (POTENTIAL).
FT CARBOHYD 577 577 N-LINKED (GLNAC...) (POTENTIAL).
FT BINDING 646 646 CHONDROITIN 4-SULFATE. CROSS-LINK SITE
FT (BY SIMILARITY).
SC SEQUENCE 886 AA. 39518 MW. AC059406852576B9 CRD64;
Query Match 6.6%; Score 94.5; DP 11; Length 886;
Best Local Similarity 23.8%; Pred. No. 14;
Matches 46; Conservative 33; Mismatches 63; Indels 19; Gaps 7;

QY 74 YVHMKRRHDFKSAEYGVQVET----LPRKDCNYKTRAEVKEPDP---LIVNRIR 124
DB 476 YFENAIL-DUTKNSYHFHYDGSSTAVAGELAPSDNNFK--ADVKGKGAINDLTETEDV 532
QY 125 LKSID--FKEGDNILSKLEYNSHNVYIMAKQKNG-KVNEKIKENLEDDGVGLADHY 182
DB 533 MKPDAAIKGGYIFGNYERLWAYTIEGLEKPKN---ANGEEKNLALAEUSLKY 589
QY 183 QGNTFGGQVLEFENHY-STOSALSKEPNE 213
DB 590 HFVTPCTPMV-KPEKN--EDQTSIAKPKGE 618

Search completed: July 16, 2003, 16:11:18
Job time : 11.6667 secs

GenCore version: 5.1.6
Copyright (c) 1993 - 2003 Compaq Inc.

OK protein - protein search, using sw model

Run on: July 16, 2003, 16:06:17 : Search time 15 seconds
(without alignments)
456,444 Million eval. operations/sec

Title: US-09-967-301-4
Perfect score: 1276
Sequence: 1 MSKCEELFTGVVPLVRLDQ3.....VLLFFVTAHHTIRKELYK 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.6

Searched: 262574 seqs, 29432922 residues

Total number of hits satisfying chosen parameters: 32/204

Minimum DB seq length: 0
Maximum DB seq length: 240000000

Post-processing: Minimum Match 95
Maximum Match 100
Listing first 25 summaries

Database : Issued_Rater's_AA*
1: /seq2.6/ptodata/1/1aa/5A_0MB.pept
2: /seq2.6/ptodata/1/1aa/5P_0MB.pept
3: /seq2.6/ptodata/1/1aa/6A_0MB.pept
4: /seq2.6/ptodata/1/1aa/6B_0MB.pept
5: /seq2.6/ptodata/1/1aa/PTMS_0MB.pept
6: /seq2.6/ptodata/1/1aa/Backlitres1.pept

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1266	99.2	236	3	US-08-893-327-16
2	1266	99.2	247	3	US-08-893-327-16
3	1262	98.9	238	1	US-08-753-143-2
4	1262	98.9	238	2	US-08-679-835-2
5	1262	98.9	238	2	US-08-680-276-2
6	1262	98.9	238	2	US-08-792-553-2
7	1262	98.9	238	3	US-08-753-144-2
8	1262	98.9	238	4	US-09-004-350-2
9	1262	98.9	238	4	US-09-172-063-2
10	1262	98.9	238	4	US-09-263-075-2
11	1262	98.9	236	4	US-08-727-452-2
12	1262	98.9	236	4	US-09-419-785-1
13	1262	98.9	1070	4	US-09-091-012A-2
14	1262	98.9	1552	4	US-09-127-227-2
15	1268	98.6	234	1	US-06-737-936-2
16	1268	98.6	234	4	US-09-121-536-1
17	1268	98.6	236	4	US-08-14007-2
18	1257	98.5	249	3	US-08-643-538-2
19	1257	98.5	249	4	US-09-503-222-2
20	1252	98.1	238	3	US-08-643-704A-49
21	1251	98.0	239	4	US-09-173-064-3
22	1251	98.0	239	4	US-09-513-783A-46
23	1251	98.0	239	4	US-09-318-929-4
24	1251	98.0	281	4	US-09-062-102-1
25	1251	98.0	281	4	US-09-354-946-1
26	1251	98.0	294	4	US-09-513-763A-2
27	1251	98.0	323	4	US-09-172-063-2

28	1251	98.0	364	4	US-09-085-305-6	Sequence 3, Appli
29	1251	98.0	459	4	US-09-513-783A-170	Sequence 170, App
30	1251	98.0	642	2	US-08-818-253-2	Sequence 2, Appli
31	1251	98.0	642	2	US-08-818-253-6	Sequence 6, Appli
32	1251	98.0	642	4	US-08-818-252-2	Sequence 2, Appli
33	1251	98.0	642	4	US-08-818-252-6	Sequence 6, Appli
34	1251	98.0	652	2	US-08-818-253-4	Sequence 4, Appli
35	1251	98.0	652	4	US-08-818-252-4	Sequence 4, Appli
36	1251	98.0	753	4	US-09-513-783A-176	Sequence 176, App
37	1251	98.0	875	4	US-09-513-783A-174	Sequence 174, App
38	1251	98.0	890	4	US-09-513-783A-174	Sequence 174, App
39	1251	98.0	941	4	US-09-513-783A-172	Sequence 172, App
40	1251	98.0	1107	4	US-06-974-543A-628	Sequence 628, App
41	1249	97.9	238	2	US-08-588-201-2	Sequence 2, Appli
42	1249	97.9	238	2	US-09-169-605-2	Sequence 2, Appli
43	1249	97.9	238	3	US-08-893-327-2	Sequence 2, Appli
44	1249	97.9	239	4	US-09-513-783A-52	Sequence 52, Appli
45	1249	97.9	247	3	US-08-893-327-20	Sequence 20, Appli

ALIGNMENTS

RESULT 1
US-08-893-327-16
Sequence 16, Application US/08093327
Patent No. 6020192
GENERAL INFORMATION:
APPLICANT: Zolotukhin, Sergei
APPLICANT: Hauswirth, William W.
APPLICANT: Mizycka, Nicholas
TITLE OF INVENTION: Humanized Green Fluorescent Protein
TITLE OF INVENTION: Genes and Methods
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arcold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,327
FILING DATE:
CLASSIFICATION: 5.4
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 9P/588,231
FILING DATE: 18-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: CFA:622\KIT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-5000
TELEFAX: (213) 789-2679
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 236 amino acids
TYPE: amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-08-893-327-16

Query Match: 99.28, Score 1246, DB 3, Length 236
Best Local Similarity: 99.28, Pred. No. 1, 66-130
Matches 236: Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSKCEELFTGVVPLVRLDQ3VSGGSGSDATYSKILKFICTGKLVPPWFTL 60

Db 1 MSKGEELTGVVPIILVELDGVNGHKFSVSSEGGDATYKGLTKPICTTKGLVFWPTL 60
Qy 61 VTTLTYGQCFSRYPDHKKRHDFFKSAMPEGYVQERTIFFKDGONKXTRAEVRFEGDLY 120
Db 61 VTTLTYGQCFSRYPDHKKRHDFFKSAMPEGYVQERTIFFKDGONKXTRAEVRFEGDLY 120
Qy 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Db 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Qy 181 HYQONTPIGDGPVLLPDNHYLSIQSALSCKDPNEKRDHNVLLFEVTAAGITRGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSIQSALSCKDPNEKRDHNVLLFEVTAAGITRGMDELYK 238
RESULT 2
US-08-893-327-18
Sequence 18, Application US/06893327
Patent No. 6020192
GENERAL INFORMATION:
APPLICANT: Zolotukhin, Sergei
APPLICANT: Haiswirth, William W.
APPLICANT: Mzyzcka, Nicholas
TITLE OF INVENTION: Mutagenized Green Fluorescent Protein
TITLE OF INVENTION: Genes and Methods
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/693,327
FILING DATE:
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/788,201
FILING DATE: 18-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,328
REFERENCE/DOCKET NUMBER: DFLA-062/KIT
TELEPHONE: (512) 418-3003
TELEFAX: (713) 743-3679
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-893-327-18
Query Match 99.2% Score 1262; DB 1; Length 238;
Best Local Similarity 99.2% Pred. No. 1.70 130;
Matches 236; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MSKGEELTGVVPIILVELDGVNGHKFSVSSEGGDATYKGLTKPICTTKGLVFWPTL 60
Db 10 MSKGEELTGVVPIILVELDGVNGHKFSVSSEGGDATYKGLTKPICTTKGLVFWPTL 60
Qy 61 VTTLTYGQCFSRYPDHKKRHDFFKSAMPEGYVQERTIFFKDGONKXTRAEVRFEGDLY 120
Db 70 VTTLTYGQCFSRYPDHKKRHDFFKSAMPEGYVQERTIFFKDGONKXTRAEVRFEGDLY 129

Qy 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Db 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Qy 181 HYQONTPIGDGPVLLPDNHYLSIQSALSCKDPNEKRDHNVLLFEVTAAGITRGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSIQSALSCKDPNEKRDHNVLLFEVTAAGITRGMDELYK 238
RESULT 3
US-08-753-143-2
Sequence 2, Application US/08753143A
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Heim, Roger
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
FILE REFERENCE: 07257/032003
CURRENT APPLICATION NUMBER: US/08/753,143A
CURRENT FILING DATE: 1996-11-20
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 2
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
US-08-753-143-2
Query Match 98.9% Score 1262; DB 1; Length 238;
Best Local Similarity 98.7% Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MSKGEELTGVVPIILVELDGVNGHKFSVSSEGGDATYKGLTKPICTTKGLVFWPTL 60
Db 1 MSKGEELTGVVPIILVELDGVNGHKFSVSSEGGDATYKGLTKPICTTKGLVFWPTL 60
Qy 61 VTTLTYGQCFSRYPDHKKRHDFFKSAMPEGYVQERTIFFKDGONKXTRAEVRFEGDLY 120
Db 61 VTTLTYGQCFSRYPDHKKRHDFFKSAMPEGYVQERTIFFKDGONKXTRAEVRFEGDLY 120
Qy 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Db 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Qy 181 HYQONTPIGDGPVLLPDNHYLSIQSALSCKDPNEKRDHNVLLFEVTAAGITRGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSIQSALSCKDPNEKRDHNVLLFEVTAAGITRGMDELYK 238
RESULT 4
US-08-679-865 2
Sequence 2, Application US/0879865
Patent No. 5912137
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Cobitt, Andrew B.
TITLE OF INVENTION: Assays for Protein Kinases Using
TITLE OF INVENTION: Fluorescent Protein Substrates
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/679,865
FILING DATE: 16-JUL-1996

5877079

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sterella, John S.
REGISTRATION NUMBER: 32,944
REFERENCE/DOCKET NUMBER: 02-072-069200
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-876-2

Query Match          98.9%   Score 1262; DB 2; Length 238;
Best Local Similarity 98.7%   Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNNGHKFSVSGEGDATYKLTAKLTCTTKLAVPWPFL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNNGHKFSVSGEGDATYKLTAKLTCTTKLAVPWPFL 60
QY 61 VTLLTYGVCCFSRYCHMKRHDFFKSAMPEGVVCERTIFFKEDSNYKTRAEVKFESGLV 120
DB 61 VTLLTYGVCCFSRYCHMKRHDFFKSAMPEGVVCERTIFFKEDSNYKTRAEVKFESGLV 120
QY 121 NRLEKLGIDFKEDGNLGHKLEYNNSHNRYIMALKQKNGIKVNEKIKENFTGGSVLAD 180
DB 121 NRLEKLGIDFKEDGNLGHKLEYNNSHNRYIMALKQKNGIKVNEKIKENFTGGSVLAD 180
QY 181 HYQONTPIGDGVLLPNNHYLSTQSALSKDPNEKKDHVLLFEFVTAAGTTRKMLHLYK 238
DB 181 HYQONTPIGDGVLLPNNHYLSTQSALSKDPNEKKDHVLLFEFVTAAGTTRKMLHLYK 238

RESULT 5
US-08-680-876-2
Sequence 2, Application: US/08520876
Patent No. 592553
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
TITLE OF INVENTION: Assays for Protein Kinases Using
TITLE OF INVENTION: Fluorescent Protein Substrates
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patattn Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/150,876
FILING DATE: 16-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sterella, John S.
REGISTRATION NUMBER: 32,944
REFERENCE/DOCKET NUMBER: 02-072-069200
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids

```

```

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-876-2

Query Match          98.9%   Score 1262; DB 2; Length 238;
Best Local Similarity 98.7%   Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNNGHKFSVSGEGDATYKLTAKLTCTTKLAVPWPFL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNNGHKFSVSGEGDATYKLTAKLTCTTKLAVPWPFL 60
QY 61 VTLLTYGVCCFSRYCHMKRHDFFKSAMPEGVVCERTIFFKEDSNYKTRAEVKFESGLV 120
DB 61 VTLLTYGVCCFSRYCHMKRHDFFKSAMPEGVVCERTIFFKEDSNYKTRAEVKFESGLV 120
QY 121 NRLEKLGIDFKEDGNLGHKLEYNNSHNRYIMALKQKNGIKVNEKIKENFTGGSVLAD 180
DB 121 NRLEKLGIDFKEDGNLGHKLEYNNSHNRYIMALKQKNGIKVNEKIKENFTGGSVLAD 180
QY 181 HYQONTPIGDGVLLPNNHYLSTQSALSKDPNEKKDHVLLFEFVTAAGTTRKMLHLYK 238
DB 181 HYQONTPIGDGVLLPNNHYLSTQSALSKDPNEKKDHVLLFEFVTAAGTTRKMLHLYK 238

RESULT 6
US-08-680-876-2
Sequence 2, Application: US/08792553
Patent No. 598100
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
TITLE OF INVENTION: Tandem Fluorescent Protein Constructs
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: FISH & RICHARDSON P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patattn Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/792,553
FILING DATE: 31-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lisa A. Hallie, Ph.D.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07257/941001/00 96-160-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5059
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-876-2

Query Match          98.9%   Score 1262; DB 2; Length 238;
Best Local Similarity 98.7%   Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNNGHKFSVSGEGDATYKLTAKLTCTTKLAVPWPFL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNNGHKFSVSGEGDATYKLTAKLTCTTKLAVPWPFL 60

```

QY 61 VTTLTGVGVCSPYDPMKRRHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 120
 DB 61 VTTSTGVGVCSPYDPMKRRHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 120
 QY 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNVTIMADKQKNGIKVNFIRHNIEASVQLAD 180
 DB 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNVTIMADKQKNGIKVNFIRHNIEASVQLAD 180
 QY 181 HYQONTPIGGGVLLPDNRHYSLSOSALKSKDPNKKRHHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 240
 DB 181 HYQONTPIGGGVLLPDNRHYSLSOSALKSKDPNKKRHHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 240

RESULT 7
 US-08-753-144-2
 ? Sequence 2: Application US/08753144
 ? Patent No. 6066476
 ? GENERAL INFORMATION:
 ? APPLICANT: Tsieng, Roger Y.
 ? APPLICANT: Holm, Roger
 ? TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
 ? FILE REFERENCE: 07257/06700
 ? CURRENT APPLICATION NUMBER: US/08/753,144
 ? CURRENT FILING DATE: 08/22/1996
 ? NUMBER OF SEQ ID NOS: 5
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Fish & Richardson, P.C.
 ? STREET: 4225 Executive Square, Suite 1400
 ? CITY: La Jolla
 ? STATE: CA
 ? COUNTRY: USA
 ? ZIP: 92037
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Diskette
 ? OPERATING SYSTEM: Windows95
 ? SOFTWARE: FastSeq for Windows Version 2.0
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/753,144
 ? FILING DATE: 08-NOV-1996
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: 08/727,452
 ? FILING DATE: 10-OCT-1996
 ? APPLICATION NUMBER: US95/14692
 ? FILING DATE: 13-NOV-1995
 ? APPLICATION NUMBER: 08/397,915
 ? FILING DATE: 13-NOV-1994
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Halpe, Lisa A.
 ? REGISTRATION NUMBER: 38,147
 ? REFERENCE/DOCKET NUMBER: 07257/06700
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 619/678-5070
 ? TELEFAX: 619/678-5099
 ? INFORMATION FOR SEQ ID NO: 2:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 238 amino acids
 ? TYPE: amino acid
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: Protein
 ? FRAGMENT TYPE: Internal
 ? US-08-753-144-2

Query Match 96.9% Score 1262; DS 4; Length 238;
 Best Local Similarity 96.7% Pred. No. 4.4e-130;
 Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNGHGFVSFGESGSDATYKLTKEICTGKLPVNPVL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNGHGFVSFGESGSDATYKLTKEICTGKLPVNPVL 60
 QY 61 VTTLTGVGVCSPYDPMKRRHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 120
 DB 61 VTTSTGVGVCSPYDPMKRRHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 120

QY 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNVTIMADKQKNGIKVNFIRHNIEASVQLAD 180
 DB 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNVTIMADKQKNGIKVNFIRHNIEASVQLAD 180
 QY 181 HYQONTPIGGGVLLPDNRHYSLSOSALKSKDPNKKRHHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 240
 DB 181 HYQONTPIGGGVLLPDNRHYSLSOSALKSKDPNKKRHHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 240

RESULT 8
 US-09-094-359-2
 ? Sequence 2: Application US/09094359
 ? Patent No. 6143132
 ? GENERAL INFORMATION:
 ? APPLICANT: Tsieng, Roger Y.
 ? APPLICANT: Miyawaki, Atsushi
 ? APPLICANT: Llopis, Juan
 ? TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
 ? MEASURING THE PH OF A BIOLOGICAL SAMPLE
 ? FILE REFERENCE: 07257/06700
 ? CURRENT APPLICATION NUMBER: US/09/094,359
 ? CURRENT FILING DATE: 1998-06-09
 ? NUMBER OF SEQ ID NOS: 18
 ? SOFTWARE: FastSeq for Windows Version 4.0
 ? SEQ ID NO 2
 ? LENGTH: 238
 ? TYPE: PRT
 ? ORGANISM: Aequorea victoria
 ? US-09-094-359-2

Query Match 96.9% Score 1262; DS 4; Length 238;
 Best Local Similarity 96.7% Pred. No. 4.4e-130;
 Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNGHGFVSFGESGSDATYKLTKEICTGKLPVNPVL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNGHGFVSFGESGSDATYKLTKEICTGKLPVNPVL 60
 QY 61 VTTLTGVGVCSPYDPMKRRHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 120
 DB 61 VTTSTGVGVCSPYDPMKRRHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 120
 QY 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNVTIMADKQKNGIKVNFIRHNIEASVQLAD 180
 DB 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNVTIMADKQKNGIKVNFIRHNIEASVQLAD 180
 QY 181 HYQONTPIGGGVLLPDNRHYSLSOSALKSKDPNKKRHHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 240
 DB 181 HYQONTPIGGGVLLPDNRHYSLSOSALKSKDPNKKRHHDFKSSAMPEGVQERTFFKCDGNKYNKTRAEVKFGDTLV 240

RESULT 9
 US-09-172-063-2
 ? Sequence 2: Application US/09172063
 ? Patent No. 6150176
 ? GENERAL INFORMATION:
 ? APPLICANT: Tsieng, Roger Y.
 ? APPLICANT: Miyawaki, Atsushi
 ? APPLICANT: Llopis, Juan
 ? APPLICANT: Wichter, Katerka M.
 ? APPLICANT: Remington, S. James
 ? TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
 ? MEASURING THE PH OF A BIOLOGICAL SAMPLE
 ? FILE REFERENCE: 07257/07100
 ? CURRENT APPLICATION NUMBER: US/09/172,063
 ? CURRENT FILING DATE: 1998-10-13
 ? EARLIER APPLICATION NUMBER: 09/094,359
 ? NUMBER OF SEQ ID NOS: 38
 ? SOFTWARE: FastSeq for Windows Version 4.0
 ? SEQ ID NO 2
 ? LENGTH: 238
 ? TYPE: PRT

```

1  ORGANISM: Aequorea victoria
US-09-172-063-2

Query Match      98.9%; Score 1262; LH 4; Length 238;
Best Local Similarity 98.7%; Pred. No. 4,4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1  MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGDATYGKLTIKFTCTGKLPVWPTL 60
DB 1  MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGDATYGKLTIKFTCTGKLPVWPTL 60

QY 61  VTITLYGVQCFSRYPDMKRRHDFKSAKPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLY 120
DB 61  VTITLYGVQCFSRYPDMKRRHDFKSAKPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLY 120

QY 61  VTITLYGVQCFSRYPDMKRRHDFKSAKPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLY 120
DB 61  VTITLYGVQCFSRYPDMKRRHDFKSAKPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLY 120

QY 121  NRIELKGIQDFKEDGNILGHKLEYNNSNVYIMAKQKNGIKVNFKIHNHEDGGVQLAD 180
DB 121  NRIELKGIQDFKEDGNILGHKLEYNNSNVYIMAKQKNGIKVNFKIHNHEDGGVQLAD 180

QY 121  NRIELKGIQDFKEDGNILGHKLEYNNSNVYIMAKQKNGIKVNFKIHNHEDGGVQLAD 180
DB 121  NRIELKGIQDFKEDGNILGHKLEYNNSNVYIMAKQKNGIKVNFKIHNHEDGGVQLAD 180

QY 181  HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNKKRDHVMVLEFVTAAGITHGMDELYK 238
DB 181  HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNKKRDHVMVLEFVTAAGITHGMDELYK 238

RESULT 10
US-09-263-975-2
: Sequence 2, Application US/09263975
: Patent No. 6248559
: GENERAL INFORMATION:
: APPLICANT: Telsio, Roger Y.
: APPLICANT: Telsio, Roger Y.
: TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR DETECTION OF ANALYTES
: FILE REFERENCE: 07257/032001
: CURRENT APPLICATION NUMBER: US/08/727,452A
: CURRENT FILING DATE: 1996-03-26
: EARLIER APPLICATION NUMBER: EPT/US95/14592
: EARLIER FILING DATE: 1995-11-13
: EARLIER APPLICATION NUMBER: US 07/337,415
: EARLIER FILING DATE: 1994-11-10
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 2
: LENGTH: 238
: TYPE: PRT
: ORGANISM: Aequorea victoria
US-09-172-063-2

Query Match      98.9%; Score 1262; LH 4; Length 238;
Best Local Similarity 98.7%; Pred. No. 4,4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1  MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGDATYGKLTIKFTCTGKLPVWPTL 60
DB 1  MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGDATYGKLTIKFTCTGKLPVWPTL 60

QY 61  VTITLYGVQCFSRYPDMKRRHDFKSAKPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLY 120
DB 61  VTITLYGVQCFSRYPDMKRRHDFKSAKPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLY 120

QY 61  VTITLYGVQCFSRYPDMKRRHDFKSAKPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLY 120
DB 61  VTITLYGVQCFSRYPDMKRRHDFKSAKPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLY 120

QY 121  NRIELKGIQDFKEDGNILGHKLEYNNSNVYIMAKQKNGIKVNFKIHNHEDGGVQLAD 180
DB 121  NRIELKGIQDFKEDGNILGHKLEYNNSNVYIMAKQKNGIKVNFKIHNHEDGGVQLAD 180

QY 121  NRIELKGIQDFKEDGNILGHKLEYNNSNVYIMAKQKNGIKVNFKIHNHEDGGVQLAD 180
DB 121  NRIELKGIQDFKEDGNILGHKLEYNNSNVYIMAKQKNGIKVNFKIHNHEDGGVQLAD 180

QY 181  HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNKKRDHVMVLEFVTAAGITHGMDELYK 238
DB 181  HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNKKRDHVMVLEFVTAAGITHGMDELYK 238

RESULT 12
US-09-418-725-2
: Sequence 1, Application US/09418785
: Patent No. 6414119
: GENERAL INFORMATION:
: APPLICANT: Fisher, Hugh
: APPLICANT: Fisher, Hugh
: TITLE OF INVENTION: Rapidly Greening, Low Oxygen Mutant of
: TITLE OF INVENTION: the Aequorea victoria Green Fluorescent Protein
: FILE REFERENCE: EPTC 99-0611
: CURRENT APPLICATION NUMBER: US/99/418,785
: CURRENT FILING DATE: 1999-10-15
: PRIOR APPLICATION NUMBER: 60/104,563
: PRIOR FILING DATE: 1998-10-16

```

```
1  NUMBER OF SEQ ID NOS: 3
2  SOFTWARE: FastSeq for Windows Version 4.0
3  SEQ ID NO 1
4  LENGTH: 238
5  TYPE: PRT
6  ORGANISM: Aquorea victoria
7  PUBLICATION INFORMATION:
8  AUTHORS: Prasher, D.C. et al.
9  TITLE: Primary structure of the Aequorea victoria green-
10  JOURNAL: Gene
11  VOLUME: 121
12  PAGES: 229-233
13  DATE: 1992-01-01
14  DATABASE ACCESSION NUMBER: Genbank M6255
15  DATABASE ENTRY DATE: 1993-04-26
16  US-09-418-785-1

Query Match          98.9%  Score 1262;  DB 4;  Length 1452;
Best Local Similarity 98.7%  Pred. No. 4 to 129;
Matches 235;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0;

QY 1 MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGSDATYKLTLPCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGSDATYKLTLPCTTCKLPVWPPTL 60
QY 61 VTITLYGVQCFSRYPDHMKRHDFPKSAMPEGVQERTIFFKDDGNKYKTRAEVKESTIV 120
DB 61 VTITLYGVQCFSRYPDHMKRHDFPKSAMPEGVQERTIFFKDDGNKYKTRAEVKESTIV 120
QY 121 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
DB 121 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238
QY 239 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
DB 239 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238

RESULT 13
US-09-351-042A-2
Sequence 2, Application US/09091342A
Patent No. 6455300
GENERAL INFORMATION:
APPLICANT: The Government of the United States of America
as represented by the Secretary
Department of Health and Human Services
Washington, D.C.
Attm Ph.D. Han
Bauer Ph.D., Gordon L.
TITLE OF INVENTION: DNA BINDING MOLECULES FOR MONITORING
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Needle & Roseberry
STREET: 127 Peachtree Street, Suite 127
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.1, Vers 3.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,042A
FILING DATE: 08-Jun-1998
CLASSIFICATION: unknown
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 62/008,473
FILING DATE: 08 Dec 1995
ATTORNEY/AGENT INFORMATION:
NAME: Selby, Elizabeth
```

```
1  REGISTRATION NUMBER: 38295
2  REFERENCE/PACKET NUMBER: 14014, 0183
3  TELECOMMUNICATION INFORMATION:
4  TELEPHONE: 404-688-6770
5  TELEFAX: 404-688-9980
6  INFORMATION FOR SEQ ID NO: 2:
7  SEQUENCE CHARACTERISTICS:
8  LENGTH: 1070 amino acids
9  TYPE: amino acid
10  TOPOLOGY: linear
11  MOLECULE TYPE: protein
12  SEQUENCE DESCRIPTION: SEQ ID NO: 2:
13  US-09-091-042A-2

Query Match          98.9%  Score 1262;  DB 4;  Length 1452;
Best Local Similarity 98.7%  Pred. No. 4 to 129;
Matches 235;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0;

QY 1 MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGSDATYKLTLPCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGSDATYKLTLPCTTCKLPVWPPTL 60
QY 61 VTITLYGVQCFSRYPDHMKRHDFPKSAMPEGVQERTIFFKDDGNKYKTRAEVKESTIV 120
DB 61 VTITLYGVQCFSRYPDHMKRHDFPKSAMPEGVQERTIFFKDDGNKYKTRAEVKESTIV 120
QY 121 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
DB 121 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238
QY 239 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
DB 239 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238

RESULT 14
US-09-127-227-2
Sequence 2, Application US/09-27227
Patent No. 6393354
GENERAL INFORMATION:
APPLICANT: David M. Kolpe
APPLICANT: Travis J. Taylor
APPLICANT: Elizabeth McNamee
TITLE OF INVENTION: Replication-Competent Virus Expressing A
TITLE OF INVENTION: Fusion Protein
FILE REFERENCE: H098-05
CURRENT APPLICATION NUMBER: US/09/127,227
CURRENT FILING DATE: 1998-07-31
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1452
TYPE: PRT
ORGANISM: herpesvirus
US-09-127-227-2

Query Match          98.9%  Score 1262;  DB 4;  Length 1452;
Best Local Similarity 98.7%  Pred. No. 6 to 129;
Matches 235;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0;

QY 1 MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGSDATYKLTLPCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLELDGVNGHKFSVSGEGSDATYKLTLPCTTCKLPVWPPTL 60
QY 61 VTITLYGVQCFSRYPDHMKRHDFPKSAMPEGVQERTIFFKDDGNKYKTRAEVKESTIV 120
DB 61 VTITLYGVQCFSRYPDHMKRHDFPKSAMPEGVQERTIFFKDDGNKYKTRAEVKESTIV 120
QY 121 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
DB 121 NRIELKGIDFKEDGNLGHKLKLYNSHNYIMADKQNGIKVNFKIRHNIEIGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNNHYLSTOSALSKDNEKRDHMLFEFVTAAGITGHMDELYK 238
```

DB 1395 HYCONTICGCPVLLPFFNHYLSTOSALSCKPNEKRDHMLFFVTAAGTTHKMTLYK 1492

RESULT IS

US-08-337-915A-2

Sequence 2, Application: US/08337915A

Patent No. 5625348

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

APPLICANT: Helix, Roger

TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Robbins, Beilner & Gausel

STREET: 201 No. 562548th Pioneerda Street, Suite 500

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90012

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.29

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/337,915A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Spitalis, John P.

REGISTRATION NUMBER: 29,275

REFERENCE/DOCKET NUMBER: 1279,139

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 947-1001

TELEFAX: (213) 977-1005

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 234 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-337-915A-2

Query Match 98.6% Score 1256 (41) Length 234

Best Local Similarity 98.5% P-Val 1.2e-124

Matches 234; Conservative 23; Mismatches 2; Gaps 0

QY 1 MSKGEELPTGVPTIVVELLDVNDKTSVSGEGDAIYSKLTAKFCTTGKLPVFAPTL 60

DB 1 MSKGEELPTGVPTIVVELLDVNDKTSVSGEGDAIYSKLTAKFCTTGKLPVFAPTL 60

QY 61 VLTYSVQCFSSVPSMKRKHDPKKSAMPDGYVOERTFFKDDNYKTRAPVKEEDITLY 120

DB 61 VLTYSVQCFSSVPSMKRKHDPKKSAMPDGYVOERTFFKDDNYKTRAPVKEEDITLY 120

QY 121 NKRIELKIGDEKEDGNLGHKLELYNNSNVYIMACKKKNQIKVNFKTHNTEKSGVLAD 180

DB 121 NKRIELKIGDEKEDGNLGHKLELYNNSNVYIMACKKKNQIKVNFKTHNTEKSGVLAD 180

QY 181 HYQONTPIGDSVLLPDHNYLSTOSALSCKPNEKRDHMLFFVTAAGTTHKMTLYK 234

DB 181 HYQONTPIGDSVLLPDHNYLSTOSALSCKPNEKRDHMLFFVTAAGTTHKMTLYK 234

QY 181 HYQONTPIGDSVLLPDHNYLSTOSALSCKPNEKRDHMLFFVTAAGTTHKMTLYK 234

DB 181 HYQONTPIGDSVLLPDHNYLSTOSALSCKPNEKRDHMLFFVTAAGTTHKMTLYK 234

Search completed: July 16, 2003, 15:15:30

Job time : 16 secs

Before version 3.1.1.6
Copyright (c) 1997 - 2003 CompuGen Ltd.

GM protein: - protein search, using sw model

Run on: July 16, 2004, 16:06:17 : Search time 15 seconds
(without alignments)
456.844 Million cell updates/sec

Title: US-09-967-301-3
Perfect score: 1276
Sequence: 1 MSKGEELFTGVVPLVLELNS.....VLQPTVIAAIIKSMKILVX 256

Scoring table: SCSM62
Gapop 10.0 , Gapext 2.5

Searched: 262574 seqs, 2942222 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 240,000,000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents, AA*
1: /cgm2_6/prodata/1/iaa/5A_0045.rep*
2: /cgm2_6/prodata/1/iaa/5A_0095.rep*
3: /cgm2_6/prodata/1/iaa/6A_0025.rep*
4: /cgm2_6/prodata/1/iaa/6A_0035.rep*
5: /cgm2_6/prodata/1/iaa/6B_0005.rep*
6: /cgm2_6/prodata/1/iaa/backfiles.rep*

Prod. No. is the number of results predicted by engine to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1258	93.6	238	1	US-08-753-144-2 Sequence 2, Appl
2	1258	93.6	238	2	US-08-679-865-2 Sequence 2, Appl
3	1258	93.6	238	2	US-08-586-876-2 Sequence 2, Appl
4	1258	93.6	238	2	US-08-792-563-2 Sequence 2, Appl
5	1258	93.6	238	3	US-08-753-144-2 Sequence 2, Appl
6	1258	93.6	238	4	US-09-094-353-2 Sequence 2, Appl
7	1258	93.6	238	4	US-09-172-063-2 Sequence 2, Appl
8	1258	93.6	238	4	US-09-263-976-2 Sequence 2, Appl
9	1258	93.6	238	4	US-09-723-432-2 Sequence 2, Appl
10	1258	93.6	238	4	US-09-418-783-1 Sequence 1, Appl
11	1255	96.4	238	3	US-08-893-327-16 Sequence 16, Appl
12	1255	96.4	247	3	US-09-893-327-16 Sequence 18, Appl
13	1254	95.3	238	1	US-08-337-315A-2 Sequence 2, Appl
14	1254	95.3	238	4	US-09-123-530-2 Sequence 2, Appl
15	1254	95.3	238	5	PCT-0595-14562-2 Sequence 2, Appl
16	1253	93.2	239	3	US-08-646-638-2 Sequence 2, Appl
17	1253	93.2	239	3	US-08-503-222-2 Sequence 2, Appl
18	1251	96.0	1070	4	US-09-091-042A-2 Sequence 2, Appl
19	1251	96.0	1452	4	US-09-127-223-2 Sequence 2, Appl
20	1248	97.8	238	3	US-08-643-734A-3 Sequence 49, Appl
21	1245	97.6	247	3	US-08-893-327-20 Sequence 20, Appl
22	1244	97.5	238	4	US-09-223-343-4 Sequence 4, Appl
23	1242	97.3	238	2	US-08-586-262-2 Sequence 2, Appl
24	1242	97.3	238	2	US-09-169-655-2 Sequence 2, Appl
25	1242	97.3	238	3	US-08-893-327-2 Sequence 2, Appl
26	1242	97.3	239	4	US-09-513-783A-2 Sequence 53, Appl
27	1242	97.3	282	4	US-09-513-783A-11 Sequence 14, Appl

26	1242	97.3	287	4	US-09-513-783A-6	Sequence 5, Appl
27	1242	97.3	291	4	US-09-513-783A-10	Sequence 10, Appl
28	1242	97.3	295	4	US-09-513-783A-16	Sequence 16, Appl
29	1242	97.3	295	4	US-09-513-783A-20	Sequence 20, Appl
30	1242	97.3	296	4	US-09-513-783A-12	Sequence 12, Appl
31	1242	97.3	302	4	US-09-513-783A-18	Sequence 18, Appl
32	1242	97.3	302	4	US-09-513-783A-26	Sequence 26, Appl
33	1242	97.3	316	4	US-09-513-783A-26	Sequence 26, Appl
34	1242	97.3	326	4	US-09-513-783A-24	Sequence 24, Appl
35	1242	97.3	347	4	US-09-513-783A-28	Sequence 28, Appl
36	1242	97.3	350	4	US-09-513-783A-30	Sequence 30, Appl
37	1242	97.3	558	4	US-09-513-783A-180	Sequence 180, Appl
38	1240	97.2	239	4	US-09-172-063-3	Sequence 3, Appl
39	1240	97.2	239	4	US-09-513-783A-46	Sequence 46, Appl
40	1240	97.2	239	4	US-09-316-919-4	Sequence 4, Appl
41	1240	97.2	241	4	US-09-062-102-1	Sequence 1, Appl
42	1240	97.2	281	4	US-09-364-946-1	Sequence 1, Appl
43	1240	97.2	284	4	US-09-513-783A-2	Sequence 2, Appl
44	1240	97.2	294	4	US-09-513-783A-2	Sequence 2, Appl
45	1240	97.2	323	4	US-09-172-063-21	Sequence 21, Appl

ALIGNMENTS

US-08-753-144-2
Sequence 2, Application: US/08753143A
GENERAL INFORMATION:
APPLICANT: Tsieng, Roger Y.
APPLICANT: Helmi, Roger
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
FILE REFERENCE: 07957/332003
CURRENT APPLICATION NUMBER: US/08753143A
CURRENT FILING DATE: 1996-11-20
NUMBER OF SEQ IDS NOS: 2
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 239
TYPE: PRT
ORGANISM: Aquorea victoria
US-08-753-144-2

Query Match	98.6%	Score	1258	DB ID	Length	239	
Best Local Similarity	98.7%	Prod. No.	5, 16-124				
Matches	245	Conservative	60	Mismatches	3	Gaps	0
QY	1	MSKGEELFTGVVPLVLELNSGSRKATATYKLTKEICTTKLQVPPPTL	60				
DB	1	MSKGEELFTGVVPLVLELNSGSRKATATYKLTKEICTTKLQVPPPTL	60				
QY	61	VTTLSGVOCPSRYFCHMKRHFERSAMPESYVOERTFEKDDGNYKTRAEVKEGTLV	120				
DB	61	VTTLSGVOCPSRYFCHMKRHFERSAMPESYVOERTFEKDDGNYKTRAEVKEGTLV	120				
QY	121	NRIELKQIDPEKEDGNILGHKLEYNKSNHNYIMADQKNGIKVNFKIHNIEDSGVOLAD	180				
DB	121	NRIELKQIDPEKEDGNILGHKLEYNKSNHNYIMADQKNGIKVNFKIHNIEDSGVOLAD	180				
QY	161	HYQONTPIGDGVLLFDNHYLSQSALSKDPNEKRDHMYLLSFTAAQITSEKELK	238				
DB	161	HYQONTPIGDGVLLFDNHYLSQSALSKDPNEKRDHMYLLSFTAAQITSEKELK	238				

US-08-679-865-2
Sequence 2, Application: US/08679865
Patent No. 5912137
GENERAL INFORMATION:
APPLICANT: Tsieng, Roger Y.
APPLICANT: Cubitt, Andrew B.
TITLE OF INVENTION: Assays for Protein Kinases Using
Fluorescent Protein Substrates
TITLE OF INVENTION: Fluorescent Protein Substrates
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:


```

1  ADDRESSEE: Townsend and Townsend and Crew LLP
2  STREET: Two Embarcadero Center, Eighth Floor
3  CITY: San Francisco
4  STATE: California
5  COUNTRY: USA
6  ZIP: 94111-3831
7  COMPUTER READABLE FORM:
8  MEDIUM TYPE: Floppy disk
9  COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: Patentin Release #1.0, Version #1.0
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/679,465
14 FILING DATE: 16-JUL-1996
15 CLASSIFICATION: 435
16 ATTORNEY/AGENT INFORMATION:
17 NAME: Storcillo, John S.
18 REGISTRATION NUMBER: 32,944
19 REFERENCE/DOCKET NUMBER: 023072-069000
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: (415) 576-0209
22 TELEFAX: (415) 576-0300
23 INFORMATION FOR SEQ ID NO: 1:
24 SEQUENCE CHARACTERISTICS:
25 LENGTH: 238 amino acids
26 TYPE: amino acid
27 TOPOLOGY: linear
28 MOLECULE TYPE: protein
29 US-08-679-855-2
30
31 Query Match: 96.6%; Score 1258; DB 2; Length 238;
32 Best Local Similarity: 96.7%; Pred. No. 5,16-129;
33 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
34
35 QY 1 MSKGEELFTGVVPIVLVDGDNVNGKFSVSGEGGDATYKLTIKFICTGKLPVWPPTL 60
36 DB 1 MSKGEELFTGVVPIVLVDGDNVNGKFSVSGEGGDATYKLTIKFICTGKLPVWPPTL 60
37
38 QY 61 VTTLSYGVOCFSRYPDHMKRHDFFKSPMPESYVQERTIFFKDCGNKTRAEYKFEEDTLY 120
39 DB 61 VTTLSYGVOCFSRYPDHMKRHDFFKSPMPESYVQERTIFFKDCGNKTRAEYKFEEDTLY 120
40
41 QY 121 NRIELKGIDKEDNGNIGLKLEYNVNSHNYYIMADKKNIGIKVNEKIRHNLEGGVOLAD 180
42 DB 121 NRIELKGIDKEDNGNIGLKLEYNVNSHNYYIMADKKNIGIKVNEKIRHNLEGGVOLAD 180
43
44 QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDPNEKRDHMLVLLGFVTAAGITHGMDELYK 238
45 DB 181 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDPNEKRDHMLVLLGFVTAAGITHGMDELYK 238
46
47 RESULT 3
48 US-08-680-876-2
49 Sequence 2, Application US/08680876
50 Patent No. 5925558
51 GENERAL INFORMATION:
52 APPLICANT: Tsien, Roger Y.
53 APPLICANT: Cubitt, Andrew B.
54 TITLE OF INVENTION: Assays for Protein Kinase Activity
55 TITLE OF INVENTION: Fluorescent Protein Substrates
56 NUMBER OF SEQUENCES: 48
57 CORRESPONDENCE ADDRESS:
58 ADDRESSEE: Townsend and Townsend and Crew LLP
59 STREET: Two Embarcadero Center, Eighth Floor
60 CITY: San Francisco
61 STATE: California
62 COUNTRY: USA
63 ZIP: 94111-3834
64 COMPUTER READABLE FORM:
65 MEDIUM TYPE: Floppy disk
66 COMPUTER: IBM PC compatible
67 OPERATING SYSTEM: PC-DOS/MS-DOS
68 SOFTWARE: Patentin Release #1.0, Version #1.0

```

```

1  CURRENT APPLICATION DATA:
2  APPLICATION NUMBER: US/08/580,676
3  FILING DATE: 16-JUL-1996
4  CLASSIFICATION: 435
5  ATTORNEY/AGENT INFORMATION:
6  NAME: Storcillo, John S.
7  REGISTRATION NUMBER: 32,944
8  REFERENCE/DOCKET NUMBER: 023072-069250
9  TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (415) 576-0200
11 TELEFAX: (415) 576-0300
12 INFORMATION FOR SEQ ID NO: 2:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 238 amino acids
15 TYPE: amino acid
16 TOPOLOGY: linear
17 MOLECULE TYPE: protein
18 US-08-680-675-2
19
20 Query Match: 96.6%; Score 1258; DB 2; Length 238;
21 Best Local Similarity: 96.7%; Pred. No. 5,16-129;
22 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
23
24 QY 1 MSKGEELFTGVVPIVLVDGDNVNGKFSVSGEGGDATYKLTIKFICTGKLPVWPPTL 60
25 DB 1 MSKGEELFTGVVPIVLVDGDNVNGKFSVSGEGGDATYKLTIKFICTGKLPVWPPTL 60
26
27 QY 61 VTTLSYGVOCFSRYPDHMKRHDFFKSPMPESYVQERTIFFKDCGNKTRAEYKFEEDTLY 120
28 DB 61 VTTLSYGVOCFSRYPDHMKRHDFFKSPMPESYVQERTIFFKDCGNKTRAEYKFEEDTLY 120
29
30 QY 121 NRIELKGIDKEDNGNIGLKLEYNVNSHNYYIMADKKNIGIKVNEKIRHNLEGGVOLAD 180
31 DB 121 NRIELKGIDKEDNGNIGLKLEYNVNSHNYYIMADKKNIGIKVNEKIRHNLEGGVOLAD 180
32
33 QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDPNEKRDHMLVLLGFVTAAGITHGMDELYK 238
34 DB 181 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDPNEKRDHMLVLLGFVTAAGITHGMDELYK 238
35
36 RESULT 4
37 US-08-792-555-2
38 Sequence 2, Application US/08792553
39 Patent No. 5941200
40 GENERAL INFORMATION:
41 APPLICANT: Tsien, Roger Y.
42 APPLICANT: Heim, Roger
43 TITLE OF INVENTION: Tandem Fluorescent Protein Constructs
44 NUMBER OF SEQUENCES: 25
45 CORRESPONDENCE ADDRESS:
46 ADDRESSEE: FISH & RICHARDSON P.C.
47 STREET: 4225 Executive Square, Suite 1400
48 CITY: La Jolla
49 STATE: California
50 COUNTRY: USA
51 ZIP: 92037
52 COMPUTER READABLE FORM:
53 MEDIUM TYPE: Floppy disk
54 COMPUTER: IBM PC compatible
55 OPERATING SYSTEM: PC-DOS/MS-DOS
56 SOFTWARE: Patentin Release #1.0, Version #1.25
57 CURRENT APPLICATION DATA:
58 APPLICATION NUMBER: US/08/792,553
59 FILING DATE: 31-JAN-1997
60 CLASSIFICATION: 435
61 ATTORNEY/AGENT INFORMATION:
62 NAME: Mesa A. Hall, Ph.D.
63 REGISTRATION NUMBER: 38,347
64 REFERENCE/DOCKET NUMBER: 07257/041001/00 96-160-2
65 TELECOMMUNICATION INFORMATION:
66 TELEPHONE: 619-678-5070
67 TELEFAX: 619-678-5099
68 INFORMATION FOR SEQ ID NO: 2:

```

```

1 SEQUENCE CHARACTERISTICS:
2 LENGTH: 238 amino acids
3 TYPE: amino acid
4 TOPOLOGY: linear
5 MOLECULE TYPE: protein
6 US-08-792-553-2

Query Match 96.6% Score 1258; DB 2; Length 238;
Best Local Similarity 98.7% Pred. No. 5,1e-129;
Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPTLVDELGVNCHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVPTLVDELGVNCHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 60
QY 61 VTLSYGVQCFSPYDPMKRRHDFKFSAMPEGYVOERTIFFKDDGNKYKRAEVKFEQDTLV 120
DB 61 VTLSYGVQCFSPYDPMKRRHDFKFSAMPEGYVOERTIFFKDDGNKYKRAEVKFEQDTLV 120
QY 121 NRLEKGDGDEKGNILGHKLEYNVNSHNYIMADKOKNGIKVNFKIRHNIEGSGVOLAD 180
DB 121 NRLEKGDGDEKGNILGHKLEYNVNSHNYIMADKOKNGIKVNFKIRHNIEGSGVOLAD 180
QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKEPNKDHVLLGFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKEPNKDHVLLGFVTAAGITHGMDELYK 238

RESULT 5
US-08-793-44-2
Sequence 2, Application US/08753:144
Patent No. 6,566,476
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Heim, Roger
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
STREET: Fish & Richardson P.C.
CITY: 4225 Executive Square, Suite 1400
STATE: La Jolla
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: PWSLSEC for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/753:144
FILING DATE: 20-NOV-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/727,452
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US95/24692
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: 05/237,915
FILING DATE: 20-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hall, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07257/032002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal

```

```

US-09-753-144-2

Query Match 96.6% Score 1258; DB 3; Length 238;
Best Local Similarity 98.7% Pred. No. 5,1e-129;
Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPTLVDELGVNCHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVPTLVDELGVNCHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 60
QY 61 VTLSYGVQCFSPYDPMKRRHDFKFSAMPEGYVOERTIFFKDDGNKYKRAEVKFEQDTLV 120
DB 61 VTLSYGVQCFSPYDPMKRRHDFKFSAMPEGYVOERTIFFKDDGNKYKRAEVKFEQDTLV 120
QY 121 NRLEKGDGDEKGNILGHKLEYNVNSHNYIMADKOKNGIKVNFKIRHNIEGSGVOLAD 180
DB 121 NRLEKGDGDEKGNILGHKLEYNVNSHNYIMADKOKNGIKVNFKIRHNIEGSGVOLAD 180
QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKEPNKDHVLLGFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKEPNKDHVLLGFVTAAGITHGMDELYK 238

RESULT 6
US-09-394-359-2
Sequence 2, Application US/09094:359
Patent No. 6,440,142
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Atsushi
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
FILE REFERENCE: 07257/067601
CURRENT APPLICATION NUMBER: US/09/394,459
CURRENT FILING DATE: 1998-06-09
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 238
TYPE: PPT
ORGANISM: Aquorea victoria
US-09-094-359-2

Query Match 96.6% Score 1258; DB 4; Length 238;
Best Local Similarity 98.7% Pred. No. 5,1e-129;
Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPTLVDELGVNCHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVPTLVDELGVNCHKFSVSGEGDATYKGLTKFKICTTGKLPVWPPTL 60
QY 61 VTLSYGVQCFSPYDPMKRRHDFKFSAMPEGYVOERTIFFKDDGNKYKRAEVKFEQDTLV 120
DB 61 VTLSYGVQCFSPYDPMKRRHDFKFSAMPEGYVOERTIFFKDDGNKYKRAEVKFEQDTLV 120
QY 121 NRLEKGDGDEKGNILGHKLEYNVNSHNYIMADKOKNGIKVNFKIRHNIEGSGVOLAD 180
DB 121 NRLEKGDGDEKGNILGHKLEYNVNSHNYIMADKOKNGIKVNFKIRHNIEGSGVOLAD 180
QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKEPNKDHVLLGFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKEPNKDHVLLGFVTAAGITHGMDELYK 238

RESULT 7
US-09-172-063-2
Sequence 2, Application US/09172:63
Patent No. 6150176
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Miyawaki, Atsushi
APPLICANT: Lloplis, Juan

```

APPLICANT: Wachter, Rebekka M.

APPLICANT: Remington, S. James

TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR DETECTION OF A PROLIFERATING PATHWAY

FILE REFERENCE: 07257/071001

CURRENT FILING DATE: 1998-10-13

EARLIER FILING DATE: 1998-06-09

NUMBER OF SEQ ID NOS: 34

SOFTWARE: FASTSEQ for Windows Version 1.0

SEQ ID NO 2

LENGTH: 238

TYPE: PRT

ORGANISM: Aequorea victoria

US-09-172-053-2

Query Match 98.6% Score 1258; DB 4; Length 238;

Best Local Similarity 98.7% Pred. No. 5,1e-129;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPTGVPELVDGVNCHRFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

DB 1 MSKGEELFTGVPTGVPELVDGVNCHRFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

QY 61 VTLSYGVOCFSRYPDHMKRHDFKSAHPGYVOERTIFFKDDGNKYKTRAEVKEFSDTLV 120

DB 61 VTLSYGVOCFSRYPDHMKRHDFKSAHPGYVOERTIFFKDDGNKYKTRAEVKEFSDTLV 120

QY 121 NRLEKGIQKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

DB 121 NRLEKGIQKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

QY 181 HYQONTPIGDPVLLPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

DB 181 HYQONTPIGDPVLLPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 8

US-09-263-975-2

Sequence 2, Application US/04264975

Patent No. 6248550

GENERAL INFORMATION:

APPLICANT: Tsieng, Roger Y.

APPLICANT: Cubitt, Andrew H.

TITLE OF INVENTION: Assays for Protein Kinases Using

TITLE OF INVENTION: Fluorescent Protein Substrates

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

ADDRESSEE: Tsieng and Townsend and Cross LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/263,975

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/479,485

FILING DATE: 14-07-1996

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John S.

REGISTRATION NUMBER: 32,844

REFERENCE/DOCKET NUMBER: 023072-065000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-263-975-2

Query Match 98.6% Score 1258; DB 4; Length 238;

Best Local Similarity 98.7% Pred. No. 5,1e-129;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPTGVPELVDGVNCHRFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

DB 1 MSKGEELFTGVPTGVPELVDGVNCHRFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

QY 61 VTLSYGVOCFSRYPDHMKRHDFKSAHPGYVOERTIFFKDDGNKYKTRAEVKEFSDTLV 120

DB 61 VTLSYGVOCFSRYPDHMKRHDFKSAHPGYVOERTIFFKDDGNKYKTRAEVKEFSDTLV 120

QY 121 NRLEKGIQKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

DB 121 NRLEKGIQKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

QY 181 HYQONTPIGDPVLLPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

DB 181 HYQONTPIGDPVLLPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 9

US-09-727-452-2

Sequence 2, Application US/08727452A

Patent No. 6319669

GENERAL INFORMATION:

APPLICANT: Tsieng, Roger Y.

APPLICANT: Helm, Roger

TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR DETECTION OF ANALYTES

FILE REFERENCE: 07257/032001

CURRENT APPLICATION NUMBER: US/08/727,452A

CURRENT FILING DATE: 1996-03-20

EARLIER APPLICATION NUMBER: PCT/US95/14692

EARLIER FILING DATE: 1995-11-13

EARLIER APPLICATION NUMBER: US 07/337,915

EARLIER FILING DATE: 1994-11-10

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FASTSEQ for Windows Version 3.0

SEQ ID NO 2

LENGTH: 238

TYPE: PRT

ORGANISM: Aequorea victoria

US-09-727-452-2

Query Match 98.6% Score 1258; DB 4; Length 238;

Best Local Similarity 98.7% Pred. No. 5,1e-129;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPTGVPELVDGVNCHRFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

DB 1 MSKGEELFTGVPTGVPELVDGVNCHRFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

QY 61 VTLSYGVOCFSRYPDHMKRHDFKSAHPGYVOERTIFFKDDGNKYKTRAEVKEFSDTLV 120

DB 61 VTLSYGVOCFSRYPDHMKRHDFKSAHPGYVOERTIFFKDDGNKYKTRAEVKEFSDTLV 120

QY 121 NRLEKGIQKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

DB 121 NRLEKGIQKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

QY 181 HYQONTPIGDPVLLPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

DB 181 HYQONTPIGDPVLLPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 10
 US-09-418-785-1
 : Sequence 1, Application US/094-8785
 : Patent No. 6414113
 : GENERAL INFORMATION:
 : APPLICANT: Fisher, Hugh
 : TITLE OF INVENTION: Rapidly Greening, Low Oxygen Mutant of
 : FILE OF INVENTION: the Aquorea victoria Green Fluorescent Protein
 : FILE REFERENCE: RUC 99-0011
 : CURRENT APPLICATION NUMBER: US/09/418,785
 : PRIOR APPLICATION NUMBER: 1999-10-15
 : PRIOR FILING DATE: 1998-10-16
 : NUMBER OF SEQ ID NOS: 3
 : SOFTWARE: FASTSEQ for Windows Version 4.0
 : SEQ ID NO 1
 : LENGTH: 238
 : TYPE: PRT
 : ORGANISM: Aquorea victoria
 : PUBCICATION INFORMATION:
 : AUTHORS: Prasher, D.C. et al.
 : TITLE: Primary structure of the Aquorea victoria Green-f
 : JOURNAL: Gene
 : VOLUME: 111
 : PAGES: 229-233
 : DATE: 1992-01-01
 : DATABASE ACCESSION NUMBER: Genbank M62553
 : DATABASE ENTRY DATE: 1993-04-26
 : US-09-418-785-1

Query Match 96.43% Score 12557 DB 47 Length 238
 Best Local Similarity 96.34% Pred. No. 116-128
 Matches 235: Conservative 0: Mismatches 3: Indels 0: Gaps 3:
 QY 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGDATYGLIKFKICTGKLVNPTL 60
 DI 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGDATYGLIKFKICTGKLVNPTL 60
 QY 51 VTLSYGVGCSFSPYDPMKRPDKSAMPDYVOERTIFFKDDGNKYKRAVKFEGSLV 120
 DB 61 VTLSYGVGCSFSPYDPMKRPDKSAMPDYVOERTIFFKDDGNKYKRAVKFEGSLV 120
 QY 121 NRLEKGDIFKEDGNIGKLEYNNSHNVYIMADKQKNGIKVNFKISHNIEDAGVOLAD 180
 DB 121 NRLEKGDIFKEDGNIGKLEYNNSHNVYIMADKQKNGIKVNFKISHNIEDAGVOLAD 180
 QY 181 HYQNTPIGDGVPILPDNHYLSTQSAISKDNKRDHNVILGFVTAAGTTHGMDELYK 238
 DB 181 HYQNTPIGDGVPILPDNHYLSTQSAISKDNKRDHNVILGFVTAAGTTHGMDELYK 238

RESULT 11
 US-08-693-327-15
 : Sequence 16, Application US/08693327
 : Patent No. 6020192
 : GENERAL INFORMATION:
 : APPLICANT: Zolotukhin, Sergei
 : APPLICANT: Hauswirth, William W.
 : APPLICANT: Mazurka, Nicholas
 : TITLE OF INVENTION: Mutarized Green Fluorescent Protein
 : TITLE OF INVENTION: Genes and Methods
 : NUMBER OF SEQUENCES: 20
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Arnold, White & Durkee
 : STREET: P. O. Box 4433
 : CITY: Houston
 : STATE: TX
 : COUNTRY: USA
 : ZIP: 77210-4433
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/893,327
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/588,201
 FILING DATE: 18-JAN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Kitchell, Barbara S.
 REGISTRATION NUMBER: 33,928
 REFERENCE/DOCKET NUMBER: UFLA:062\KIT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3010
 TELEFAX: (713) 789-2679
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 239 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: Protein
 US-08-893-327-16

Query Match 96.49% Score 12557 DB 37 Length 238
 Best Local Similarity 96.34% Pred. No. 116-128
 Matches 234: Conservative 2: Mismatches 2: Indels 0: Gaps 0:
 QY 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGDATYGLIKFKICTGKLVNPTL 60
 DI 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGDATYGLIKFKICTGKLVNPTL 60
 QY 51 VTLSYGVGCSFSPYDPMKRPDKSAMPDYVOERTIFFKDDGNKYKRAVKFEGSLV 120
 DB 51 VTLSYGVGCSFSPYDPMKRPDKSAMPDYVOERTIFFKDDGNKYKRAVKFEGSLV 120
 QY 121 NRLEKGDIFKEDGNIGKLEYNNSHNVYIMADKQKNGIKVNFKISHNIEDAGVOLAD 180
 DB 121 NRLEKGDIFKEDGNIGKLEYNNSHNVYIMADKQKNGIKVNFKISHNIEDAGVOLAD 180
 QY 181 HYQNTPIGDGVPILPDNHYLSTQSAISKDNKRDHNVILGFVTAAGTTHGMDELYK 238
 DB 181 HYQNTPIGDGVPILPDNHYLSTQSAISKDNKRDHNVILGFVTAAGTTHGMDELYK 238

RESULT 14
 US-08-693-327-13
 : Sequence 16, Application US/08693327
 : Patent No. 6020192
 : GENERAL INFORMATION:
 : APPLICANT: Zolotukhin, Sergei
 : APPLICANT: Hauswirth, William W.
 : APPLICANT: Mazurka, Nicholas
 : TITLE OF INVENTION: Mutarized Green Fluorescent Protein
 : TITLE OF INVENTION: Genes and Methods
 : NUMBER OF SEQUENCES: 20
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Arnold, White & Durkee
 : STREET: P. O. Box 4433
 : CITY: Houston
 : STATE: TX
 : COUNTRY: USA
 : ZIP: 77210-4433
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent in Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/893,327
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/586,201
FILING DATE: 18-CAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara E.
REGISTRATION NUMBER: 24,528
REFERENCE/DOCKET NUMBER: 01LA0562AKIT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-0060
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-693-327-19

Query Match 98.3% Score 1254; DB 1; Length 238;
Best Local Similarity 98.3%; Pred. No. 1.4e-128;
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVWPLVLELDGVNNGHKFSVSGEGGATYCKLTLPKPICTISKLPPWPTL 60
DB 1 MSKGEELFTGVWPLVLELDGVNNGHKFSVSGEGGATYCKLTLPKPICTISKLPPWPTL 60
QY 61 VTTLISYGVQCFSEYPOHMKRHDFFKSAPEGVYQERTFFKDGNGYKRAEYKFEEDLY 120
DB 61 VTTLISYGVQCFSEYPOHMKRHDFFKSAPEGVYQERTFFKDGNGYKRAEYKFEEDLY 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEGSGVQLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEGSGVQLAD 180
QY 181 HYQONTPIGDGVPVLLPNDNYLSTQSALSKEPNKRDHNVLLFVTAAGTTHRMDELYK 238
DB 181 HYQONTPIGDGVPVLLPNDNYLSTQSALSKEPNKRDHNVLLFVTAAGTTHRMDELYK 238

RESULT 13
US-08-337-915A-2
Sequence 2, Application US/08337915A
Patent No. 5625048
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Tsukamoto, Toshio
APPLICANT: Yamasaki, Masao-Oshi
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reddus, Berliner & Carson
STREET: 201 NO. 562504th Figueroa Street, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90012
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/337,915A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spitals, John P.
REGISTRATION NUMBER: 29,215
REFERENCE/DOCKET NUMBER: 1279-178
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 977-1001
TELEFAX: (213) 977-1003
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids

```

```

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-121-539-1

Query Match 98.3% Score 1254; DB 1; Length 238;
Best Local Similarity 98.3%; Pred. No. 1.4e-128;
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVWPLVLELDGVNNGHKFSVSGEGGATYCKLTLPKPICTISKLPPWPTL 60
DB 1 MSKGEELFTGVWPLVLELDGVNNGHKFSVSGEGGATYCKLTLPKPICTISKLPPWPTL 60
QY 61 VTTLISYGVQCFSEYPOHMKRHDFFKSAPEGVYQERTFFKDGNGYKRAEYKFEEDLY 120
DB 61 VTTLISYGVQCFSEYPOHMKRHDFFKSAPEGVYQERTFFKDGNGYKRAEYKFEEDLY 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEGSGVQLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEGSGVQLAD 180
QY 181 HYQONTPIGDGVPVLLPNDNYLSTQSALSKEPNKRDHNVLLFVTAAGTTHRMDELYK 238
DB 181 HYQONTPIGDGVPVLLPNDNYLSTQSALSKEPNKRDHNVLLFVTAAGTTHRMDELYK 238

RESULT 14
US-09-121-539-1
Sequence 1, Application US/09121539B
Patent No. 6155416
GENERAL INFORMATION:
APPLICANT: Osumi, Takashi
APPLICANT: Tsukamoto, Toshio
APPLICANT: Yamasaki, Masao-Oshi
TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reddus, Berliner & Carson
STREET: 201 NO. 562504th Figueroa Street, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90012
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/121,539B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spitals, John P.
REGISTRATION NUMBER: 29,215
REFERENCE/DOCKET NUMBER: 1279-178
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 977-1001
TELEFAX: (213) 977-1003
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids

```

RESULT 15
PCT-US95-14692-2
Sequence 2: Application: PC/TOS9514692
GENERAL INFORMATION:
APPLICANT: Tsieng, Roger Y.
APPLICANT: Tsieng, Roger
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Robbins, Berliner & Carsen
STREET: 201 North Figueroa Street, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90012
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent's Release #1.0, Version #1.15
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14692
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Spittals, John P.
REGISTRATION NUMBER: 29,215
REFERENCE/DOCKET NUMBER: 1279-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 977-1001
TELEFAX: (213) 977-1001
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amio acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-14692-2

Query Match 98.3% Score 1244; DB: 5; Date: 212
Best Local Similarity 98.3% Prod. No: 1.40-124
Matches 234; Conservative 1; Misses 1; Gaps 0
CY 1 MSKGEELPTGVVPIVDELQDVNGHREFSVSGESFATVCKTILAFDTETKLPVWETI 60
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
FH 1 MSKGEELPTGVVPIVDELQDVNGHREFSVSGESFATVCKTILAFDTETKLPVWETI 60
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
CY 61 VTTLVGVGVCPSSYKGMKKAHIFPKSAMTEVYVCEFTTEKQKQNKYKTAQKSEFEDLY 120
III IIII IIII IIII IIII IIII IIII IIII IIII
DA 61 VTTFSGVGVCPSSYKGMKKAHIFPKSAMTEVYVCEFTTEKQKQNKYKTAQKSEFEDLY 120
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
CY 121 NRLELKTETPKDGNLGRKLEYNYSNHVYIMAKKQKNIKYNKIKRNIETGNVCLAG 180
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
DA 121 NRLELKTETPKDGNLGRKLEYNYSNHVYIMAKKQKNIKYNKIKRNIETGNVCLAG 180
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
CY 181 HYQNTPIGPGVLLDNNHVLSTQSALSKEPKKRMVLLFVYIAGTIHSMDEK 238
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
DA 181 HYQNTPIGPGVLLDNNHVLSTQSALSKEPKKRMVLLFVYIAGTIHSMDEK 238
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII

Search completed: July 16, 2003, 16:15:29
Job time : 16 secs

ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/579,445
 FILING DATE: 16-JUL-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Storella, John S.
 REGISTRATION NUMBER: 32,944
 REFERENCE/DOCKET NUMBER: 024072-069200
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-679-865-2

Query Match 100.0% Score 1275: 98.2% Length 238:
 Best Local Similarity 100.0%: Pred No. 1e-129:
 Matches 238: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
 QY 1 MSKGEELFTGVVFLVELDGVNKHKFSVSGEGDATYVKLTIKFICTGKLVKPTL 60
 DB 1 MSKGEELFTGVVFLVELDGVNKHKFSVSGEGDATYVKLTIKFICTGKLVKPTL 60
 QY 61 VTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDGNSKTRAEVKEGTLV 120
 DB 61 VTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDGNSKTRAEVKEGTLV 120
 QY 121 NRLEKIDFKEDGNILGKLEYNYNHNVYIMADKOKNIGKVNKIRHNIEGSGVLAD 180
 DB 121 NRLEKIDFKEDGNILGKLEYNYNHNVYIMADKOKNIGKVNKIRHNIEGSGVLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITHGMKIELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITHGMKIELYK 238

RESULT 3
 US-08-580-876-2
 Sequence 2, Application US/28683876
 Patent No. 5925558
 GENERAL INFORMATION:
 APPLICANT: Isien, Roger Y.
 APPLICANT: Cubitt, Andrew B.
 TITLE OF INVENTION: Assays for Protein Kinases Using
 TITLE OF INVENTION: Fluorescent Protein Substrates
 NUMBER OF SEQUENCES: 49
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.0

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/580,876
 FILING DATE: 16-JUL-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Storella, John S.
 REGISTRATION NUMBER: 32,944
 REFERENCE/DOCKET NUMBER: 023072-069200
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-679-865-2
 Query Match 100.0% Score 1275: 98.2% Length 238:
 Best Local Similarity 100.0%: Pred No. 1e-129:
 Matches 238: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
 QY 1 MSKGEELFTGVVFLVELDGVNKHKFSVSGEGDATYVKLTIKFICTGKLVKPTL 60
 DB 1 MSKGEELFTGVVFLVELDGVNKHKFSVSGEGDATYVKLTIKFICTGKLVKPTL 60
 QY 61 VTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDGNSKTRAEVKEGTLV 120
 DB 61 VTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDGNSKTRAEVKEGTLV 120
 QY 121 NRLEKIDFKEDGNILGKLEYNYNHNVYIMADKOKNIGKVNKIRHNIEGSGVLAD 180
 DB 121 NRLEKIDFKEDGNILGKLEYNYNHNVYIMADKOKNIGKVNKIRHNIEGSGVLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITHGMKIELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITHGMKIELYK 238

RESULT 4
 US-08-792-553-2
 Sequence 2, Application US/08792553
 Patent No. 5981200
 GENERAL INFORMATION:
 APPLICANT: Isien, Roger Y.
 APPLICANT: Helm, Roger
 TITLE OF INVENTION: Tandem Fluorescent Protein Constructs
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FISH & RICHARDSON P.C.
 STREET: 4225 Executive Square, Suite 1400
 CITY: La Jolla
 STATE: California
 COUNTRY: USA
 ZIP: 92037
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/792,553
 FILING DATE: 31-JAN-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Lisa A. Hallie, Ph.D.
 REGISTRATION NUMBER: 32,347
 REFERENCE/DOCKET NUMBER: 07257/041001/UC 96-150-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-678-5070
 TELEFAX: 619-678-5049
 INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-09-792-553-2

Query Match 100.0% Score 1275; DB 2; Length 238;
 Best Local Similarity 100.0% Pred. No. 10-129;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPLVDELGDVNGHKFSVSGEGSDATYKTLKFTCTTGKLPVWFTL 60
 DB 1 MSKGEELFTGVVPLVDELGDVNGHKFSVSGEGSDATYKTLKFTCTTGKLPVWFTL 60
 QY 61 VTTFSYGVQCFESRYPDHMKRHDFKFSAMPYGVQERTIFFKDGNGYKTRAEYKFSGLTV 120
 DB 61 VTTFSYGVQCFESRYPDHMKRHDFKFSAMPYGVQERTIFFKDGNGYKTRAEYKFSGLTV 120
 QY 121 NRIELKGIDFKEDGNLGHKLEYNINSHNYIMADKQKNGIKYVNFKIRHNIEGSSVLAD 180
 DB 121 NRIELKGIDFKEDGNLGHKLEYNINSHNYIMADKQKNGIKYVNFKIRHNIEGSSVLAD 180
 QY 181 HYQONTPIGDGPFVLLPDNHYLSTQSALSQKPNKRDHMHVLEFVTAAGTTHGMCELYK 238
 DB 181 HYQONTPIGDGPFVLLPDNHYLSTQSALSQKPNKRDHMHVLEFVTAAGTTHGMCELYK 238

RESULT 5

US-09-753-144-2
 Sequence 2, Application US/08753144
 Patent No. 606476

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.
 APPLICANT: He, Roger
 TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Elson A. Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM

MEDIUM TYPE: Diskette

COMPIER: IBM Compatible

OPERATING SYSTEM: Windows95

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08753,144

FILING DATE: 20-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/727,452

FILING DATE: 10-OCT-1996

APPLICATION NUMBER: US95/14692

FILING DATE: 13-NOV-1995

APPLICATION NUMBER: 08/337,915

FILING DATE: 10-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Halls, Lisa A.

REGISTRATION NUMBER: 38,147

REFERENCE/POCKET NUMBER: C7257/C32002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/478-5070

TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

US-09-753-144-2

Query Match 100.0% Score 1275; DB 3; Length 238;
 Best Local Similarity 100.0% Pred. No. 10-129;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPLVDELGDVNGHKFSVSGEGSDATYKTLKFTCTTGKLPVWFTL 60
 DB 1 MSKGEELFTGVVPLVDELGDVNGHKFSVSGEGSDATYKTLKFTCTTGKLPVWFTL 60
 QY 61 VTTFSYGVQCFESRYPDHMKRHDFKFSAMPYGVQERTIFFKDGNGYKTRAEYKFSGLTV 120
 DB 61 VTTFSYGVQCFESRYPDHMKRHDFKFSAMPYGVQERTIFFKDGNGYKTRAEYKFSGLTV 120
 QY 121 NRIELKGIDFKEDGNLGHKLEYNINSHNYIMADKQKNGIKYVNFKIRHNIEGSSVLAD 180
 DB 121 NRIELKGIDFKEDGNLGHKLEYNINSHNYIMADKQKNGIKYVNFKIRHNIEGSSVLAD 180
 QY 181 HYQONTPIGDGPFVLLPDNHYLSTQSALSQKPNKRDHMHVLEFVTAAGTTHGMCELYK 238
 DB 181 HYQONTPIGDGPFVLLPDNHYLSTQSALSQKPNKRDHMHVLEFVTAAGTTHGMCELYK 238

RESULT 6

US-09-964-359-2
 Sequence 2, Application US/09094359
 Patent No. 614032

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

APPLICANT: Miyawaki, Atsushi

APPLICANT: Clopis, Juan

TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR

FILE REFERENCE: 07257/067003

CURRENT APPLICATION NUMBER: US/99/094,359

CURRENT FILING DATE: 1998-06-09

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Vers.00.4.0

SEQ ID NO: 2

LENGTH: 234

TYPE: PRT

ORGANISM: Anquorea victoria

US-09-964-359-2

Query Match 100.0% Score 1275; DB 4; Length 238;
 Best Local Similarity 100.0% Pred. No. 10-129;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPLVDELGDVNGHKFSVSGEGSDATYKTLKFTCTTGKLPVWFTL 60
 DB 1 MSKGEELFTGVVPLVDELGDVNGHKFSVSGEGSDATYKTLKFTCTTGKLPVWFTL 60
 QY 61 VTTFSYGVQCFESRYPDHMKRHDFKFSAMPYGVQERTIFFKDGNGYKTRAEYKFSGLTV 120
 DB 61 VTTFSYGVQCFESRYPDHMKRHDFKFSAMPYGVQERTIFFKDGNGYKTRAEYKFSGLTV 120
 QY 121 NRIELKGIDFKEDGNLGHKLEYNINSHNYIMADKQKNGIKYVNFKIRHNIEGSSVLAD 180
 DB 121 NRIELKGIDFKEDGNLGHKLEYNINSHNYIMADKQKNGIKYVNFKIRHNIEGSSVLAD 180
 QY 181 HYQONTPIGDGPFVLLPDNHYLSTQSALSQKPNKRDHMHVLEFVTAAGTTHGMCELYK 238
 DB 181 HYQONTPIGDGPFVLLPDNHYLSTQSALSQKPNKRDHMHVLEFVTAAGTTHGMCELYK 238

RESULT 7

US-09-172-063-2
 Sequence 2, Application US/99172063
 Patent No. 6150176

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

APPLICANT: Miyawaki, Atsushi

APPLICANT: Clopis, Juan

```

7 APPLICANT: Wachter, Rebecca M.
7 APPLICANT: Richardson, S. James
7 TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
7 TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
7 FILE REFERENCE: 07257/071001
7 CURRENT APPLICATION NUMBER: US/09/272,163
7 CURRENT FILING DATE: 1998-10-13
7 EARLIER APPLICATION NUMBER: 07/094,459
7 EARLIER FILING DATE: 1998-06-03
7 NUMBER OF SEQ ID NOS: 38
7 SOFTWARE: FastSeq for Windows Version 4.0
7 SEQ ID NO 2
7 LENGTH: 238
7 TYPE: PRT
7 ORGANISM: Aequorea victoria
7 US-09-172-063-2

Query Match 100.0% Score 1275; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1e-129;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPILVELDGVNCHKFSVSGEGDATYCKLTKLFICTGKLPVWPPTL 60
DB 1 MSKGEELFTGVPILVELDGVNCHKFSVSGEGDATYCKLTKLFICTGKLPVWPPTL 60
QY 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKFEEDTLV 120
DB 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKFEEDTLV 120
QY 121 NRLEKGIQDKFKEDGNLGHKLEYNYNHSHVYIMADKQKNGIKVNFKIRHNIEGSSVQLAD 180
DB 121 NRLEKGIQDKFKEDGNLGHKLEYNYNHSHVYIMADKQKNGIKVNFKIRHNIEGSSVQLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238

RESULT 5
US-08-727-452-2
7 Sequence 2: Application US/08/27452A
7 Patent No. 639669
7 GENERAL INFORMATION:
7 APPLICANT: Tsiel, Roger Y.
7 APPLICANT: Helm, Roger
7 TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR DETECTION OF ANALYTES
7 FILE REFERENCE: 07257/032031
7 CURRENT APPLICATION NUMBER: US/08/727,452A
7 EARLIER FILING DATE: 1996-03-20
7 EARLIER APPLICATION NUMBER: PCT/US95/14692
7 EARLIER FILING DATE: 1995-11-13
7 EARLIER APPLICATION NUMBER: US 07/337,915
7 EARLIER FILING DATE: 1994-11-10
7 NUMBER OF SEQ ID NOS: 5
7 SOFTWARE: FastSeq for Windows Version 3.0
7 SEQ ID NO 2
7 LENGTH: 239
7 TYPE: PRT
7 ORGANISM: Aequorea victoria
7 US-08-727-452-2

Query Match 100.0% Score 1275; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1e-129;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPILVELDGVNCHKFSVSGEGDATYCKLTKLFICTGKLPVWPPTL 60
DB 1 MSKGEELFTGVPILVELDGVNCHKFSVSGEGDATYCKLTKLFICTGKLPVWPPTL 60
QY 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKFEEDTLV 120
DB 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKFEEDTLV 120
QY 121 NRLEKGIQDKFKEDGNLGHKLEYNYNHSHVYIMADKQKNGIKVNFKIRHNIEGSSVQLAD 180
DB 121 NRLEKGIQDKFKEDGNLGHKLEYNYNHSHVYIMADKQKNGIKVNFKIRHNIEGSSVQLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238

APPLICANT: Wachter, Rebecca M.
APPLICANT: Richardson, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/09/272,163
CURRENT FILING DATE: 1998-10-13
EARLIER APPLICATION NUMBER: 07/094,459
EARLIER FILING DATE: 1998-06-03
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-172-063-2

Query Match 100.0% Score 1275; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1e-129;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPILVELDGVNCHKFSVSGEGDATYCKLTKLFICTGKLPVWPPTL 60
DB 1 MSKGEELFTGVPILVELDGVNCHKFSVSGEGDATYCKLTKLFICTGKLPVWPPTL 60
QY 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKFEEDTLV 120
DB 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKFEEDTLV 120
QY 121 NRLEKGIQDKFKEDGNLGHKLEYNYNHSHVYIMADKQKNGIKVNFKIRHNIEGSSVQLAD 180
DB 121 NRLEKGIQDKFKEDGNLGHKLEYNYNHSHVYIMADKQKNGIKVNFKIRHNIEGSSVQLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238

RESULT 6
US-09-263-975-2
7 Sequence 2: Application US/09/263975
7 Patent No. 6248550
7 GENERAL INFORMATION:
7 APPLICANT: Tsiel, Roger Y.
7 APPLICANT: Cubitt, Andrew B.
7 TITLE OF INVENTION: Assays for Protein Kinases Using
7 TITLE OF INVENTION: Fluorescent Protein Substrates
7 NUMBER OF SEQUENCES: 48
7 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Townsend and Townsend and Crew LLP
7 STREET: Two Embarcadero Center, Eighth Floor
7 CITY: San Francisco
7 STATE: California
7 COUNTRY: USA
7 ZIP: 94111-3834
7 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM PC compatible
7 OPERATING SYSTEM: PC-DOS/MS-DOS
7 SOFTWARE: Patent In Release 4.0.0, Version 4.0.0
7 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/09/263,975
7 FILING DATE:
7 CLASSIFICATION:
7 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: US 08/639,845
7 FILING DATE: 12-JUL-1994
7 ATTORNEY/AGENT INFORMATION:
7 NAME: Stereella, John S.
7 REGISTRATION NUMBER: 32,444
7 REFERENCE/DOCKET NUMBER: 023072-064000
7 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: (415) 576-0260

```

```

RESULT 30
US-09-418-785-1
? Sequence 1, Application US/09418785
? Patent No. 6414119
? GENERAL INFORMATION:
? APPLICANT: Fisher, Ruth
? TITLE OF INVENTION: Rapidly Greening, Low Oxygen Mutant of
? TITLE OF INVENTION: The Aequorea victoria Green Fluorescent Protein
? FILE REFERENCE: Pats 99-5011
? CURRENT APPLICATION NUMBER: US/09/418,785
? CURRENT FILING DATE: 1999-10-15
? PRIOR APPLICATION NUMBER: 60/104,563
? PRIOR FILING DATE: 1998-10-16
? NUMBER OF SEQ ID NOS: 3
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 1
? LENGTH: 238
? TYPE: PR
? ORGANISM: Aequorea victoria
? PUBLICATION INFORMATION:
? AUTHORS: Prasher, D.C. et al.
? TITLE: Primary structure of the Aequorea victoria green-1
? JOURNAL: Gene
? VOLUME: 211
? PAGES: 229-233
? DATE: 1995-01-01
? DATABASE ACCESSION NUMBER: Genbank M6265+
? DATABASE ENTRY DATE: 1993-04-26
US-09-418-785-1

```

```

Query Match          100.0%   Score 1275   DB 1: Length 238
Best Local Similarity 100.0%   Pred. No. 2.7e-129
Matches 238: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYGKLTLSKFLICITGKLPVNPRL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYGKLTLSKFLICITGKLPVNPRL 60

QY 61 VTTFSGVGVCFSPYDMKRRHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEEDTLV 120
DB 61 VTTFSGVGVCFSPYDMKRRHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEEDTLV 120

QY 121 NRLEKGDIDFKEDGNILGHKLEYNINSVNYIMADKKNGIKVAEKIRHNIHGSGVQLAD 180
DB 121 NRLEKGDIDFKEDGNILGHKLEYNINSVNYIMADKKNGIKVAEKIRHNIHGSGVQLAD 180

QY 181 HYQONTPIGDSPVLPDNYHLYSTQSAISKDPNEKRDHNVLEFFVTAAGITHGMPELYK 238
DB 181 HYQONTPIGDSPVLPDNYHLYSTQSAISKDPNEKRDHNVLEFFVTAAGITHGMPELYK 238

```

```

RESULT 31
US-09-337-915A-2
? Sequence 2, Application US/09337915A
? Patent No. 5625348
? GENERAL INFORMATION:
? APPLICANT: Isien, Roger Y.
? APPLICANT: Genm, Takashi
? TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Robbins, Berliner & Carson
? STREET: 201 No. 5625048th Figueroa Street, Suite 500
? CITY: Los Angeles
? STATE: California
? COUNTRY: USA
? ZIP: 90012
? COMPUTER READERABLE FORM:
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version: #1.25

```

```

? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/337,915A
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Spitals, John P.
? REGISTRATION NUMBER: 29,215
? REFERENCE/DOCKET NUMBER: 1279-178
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (213) 977-1001
? TELEFAX: (213) 977-1003
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 238 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-09-337-915A-2

```

```

Query Match          99.7%   Score 1271   DB 1: Length 238
Best Local Similarity 99.6%   Pred. No. 2.7e-129
Matches 237: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYGKLTLSKFLICITGKLPVNPRL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYGKLTLSKFLICITGKLPVNPRL 60

QY 61 VTTFSGVGVCFSPYDMKRRHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEEDTLV 120
DB 61 VTTFSGVGVCFSPYDMKRRHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEEDTLV 120

QY 121 NRLEKGDIDFKEDGNILGHKLEYNINSVNYIMADKKNGIKVAEKIRHNIHGSGVQLAD 180
DB 121 NRLEKGDIDFKEDGNILGHKLEYNINSVNYIMADKKNGIKVAEKIRHNIHGSGVQLAD 180

QY 181 HYQONTPIGDSPVLPDNYHLYSTQSAISKDPNEKRDHNVLEFFVTAAGITHGMPELYK 238
DB 181 HYQONTPIGDSPVLPDNYHLYSTQSAISKDPNEKRDHNVLEFFVTAAGITHGMPELYK 238

```

```

RESULT 12
US-09-121-539-1
? Sequence 1, Application US/091215395
? Patent No. 6194548
? GENERAL INFORMATION:
? APPLICANT: Genm, Takashi
? APPLICANT: Tsuchimoto, Toshiro
? APPLICANT: Tsuchimoto, No. 5194548
? APPLICANT: Yamashita, Masayoshi
? TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
? TITLE OF INVENTION: PROTEINS
? FILE REFERENCE: 046124-5035
? CURRENT APPLICATION NUMBER: US/09/121,539
? CURRENT FILING DATE: 1998-07-24
? PRIOR APPLICATION NUMBER: JP 025418/1998
? PRIOR FILING DATE: 1998-01-24
? NUMBER OF SEQ ID NOS: 14
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1
? LENGTH: 236
? TYPE: PR
? ORGANISM: Aequorea victoria
? FEATURE:
? OTHER INFORMATION: Green fluorescent protein
US-09-121-539-1

```

```

Query Match          99.7%   Score 1271   DB 4: Length 238
Best Local Similarity 99.6%   Pred. No. 2.7e-129
Matches 237: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYGKLTLSKFLICITGKLPVNPRL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYGKLTLSKFLICITGKLPVNPRL 60

```

```

QY 61 VTTSYGVGCFSPY:DHMKKRDFFKSAPEGVGCHPTFFKDDGNKTRAEVKEGDTLVN 121
DB 61 VTTSYGVGCFSPY:DHMKKRDFFKSAPEGVGCHPTFFKDDGNKTRAEVKEGDTLVN 121
QY 121 NRTELKGDIFKEDGNLQDK:LYNNYNHNVYIMAKKKNK:KVNKKTRHN:LEFVS:LAAD 121
DB 121 NRTELKGDIFKEDGNLQDK:LYNNYNHNVYIMAKKKNK:KVNKKTRHN:LEFVS:LAAD 121
QY 181 HYQONTPIGDSVLLPDNNH:STOSALS:KDNKRDHNV:LEFVTAAGITH:MEELK 238
DB 181 HYQONTPIGDSVLLPDNNH:STOSALS:KDNKRDHNV:LEFVTAAGITH:MEELK 238

RESULT 13
PCT-US95-14692-2
: Sequence 2, Application PC/TUS9514692
: GENERAL INFORMATION:
: APPLICANT: Helim, Roger Y.
: TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Robbins, Berlin & Carlson
: STREET: 201 North Figueroa Street, Suite 500
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90012
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/14692
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Spitals, John P.
: REGISTRATION NUMBER: 29,215
: REFERENCE/DOCKET NUMBER: 1279-178
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 977-1001
: TELEFAX: (213) 977-1003
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 238 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US95-14692-2

Query Match 99.78; Score 1271; DB 5; Length 238;
Best Local Similarity 99.68; Pred. No. 2,76-179;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGSDATYVKLTIKFICTTCKLPVWPTLV 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGSDATYVKLTIKFICTTCKLPVWPTLV 60
QY 61 VTTSYGVGCFSPY:DHMKKRDFFKSAPEGVGCHPTFFKDDGNKTRAEVKEGDTLVN 120
DB 61 VTTSYGVGCFSPY:DHMKKRDFFKSAPEGVGCHPTFFKDDGNKTRAEVKEGDTLVN 120
QY 121 NRTELKGDIFKEDGNLQDK:LYNNYNHNVYIMAKKKNK:KVNKKTRHN:LEFVS:LAAD 121
DB 121 NRTELKGDIFKEDGNLQDK:LYNNYNHNVYIMAKKKNK:KVNKKTRHN:LEFVS:LAAD 121
QY 181 HYQONTPIGDSVLLPDNNH:STOSALS:KDNKRDHNV:LEFVTAAGITH:MEELK 238
DB 181 HYQONTPIGDSVLLPDNNH:STOSALS:KDNKRDHNV:LEFVTAAGITH:MEELK 238

```

```

RESULT 14
US-08-646-536-2
: Sequence 2, Application US/08646536
: Patent No. 6027681
: GENERAL INFORMATION:
: APPLICANT: Pavlakis, George N.
: APPLICANT: Galanaris, George A.
: APPLICANT: Stauber, Roland H.
: APPLICANT: Vourdakis, John N.
: TITLE OF INVENTION: Mutant Aequorea victoria Fluorescent
: PROTEINS Having Increased Cellular Fluorescence
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, 5th Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/646,538
: FILING DATE: NO. 6027681 yet assigned
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 015280 249000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 239 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-646-536-2

Query Match 99.68; Score 1270; DB 3; Length 239;
Best Local Similarity 100.0%; Pred. No. 3 5e-129;
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGSDATYVKLTIKFICTTCKLPVWPTLV 61
DB 2 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGSDATYVKLTIKFICTTCKLPVWPTLV 61
QY 3 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGSDATYVKLTIKFICTTCKLPVWPTLV 62
DB 3 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGSDATYVKLTIKFICTTCKLPVWPTLV 62
QY 62 TTTSYGVGCFSPY:DHMKKRDFFKSAPEGVGCHPTFFKDDGNKTRAEVKEGDTLVN 121
DB 62 TTTSYGVGCFSPY:DHMKKRDFFKSAPEGVGCHPTFFKDDGNKTRAEVKEGDTLVN 121
QY 63 TTTSYGVGCFSPY:DHMKKRDFFKSAPEGVGCHPTFFKDDGNKTRAEVKEGDTLVN 122
DB 63 TTTSYGVGCFSPY:DHMKKRDFFKSAPEGVGCHPTFFKDDGNKTRAEVKEGDTLVN 122
QY 121 NRTELKGDIFKEDGNLQDK:LYNNYNHNVYIMAKKKNK:KVNKKTRHN:LEFVS:LAAD 161
DB 121 NRTELKGDIFKEDGNLQDK:LYNNYNHNVYIMAKKKNK:KVNKKTRHN:LEFVS:LAAD 162
QY 182 HYQONTPIGDSVLLPDNNH:STOSALS:KDNKRDHNV:LEFVTAAGITH:MEELK 238
DB 182 HYQONTPIGDSVLLPDNNH:STOSALS:KDNKRDHNV:LEFVTAAGITH:MEELK 239

RESULT 15
US-08-539-222-2
: Sequence 2, Application US/08503222
: Patent No. 6265548
: GENERAL INFORMATION:
: APPLICANT: Pavlakis, George N.
: APPLICANT: Galanaris, George A.
: APPLICANT: Stauber, Roland H.

```

1 APPLICANT: Vedinakis, John N.
 2 TITLE OF INVENTION: Mutant Acquireda Virusin Fluorococor
 3 TITLE OF INVENTION: Proteins Having Increased Cellular Fl uorescence
 4 NUMBER OF SEQUENCES: 17
 5 CORRESPONDENCE ADDRESS:
 6 ADDRESSEE: Townsend and Townsend and Crew LLP
 7 STREET: Two Embarcadero Center, 8th Floor
 8 CITY: San Francisco
 9 STATE: California
 10 COUNTRY: USA
 11 ZIP: 94111-3834
 12 COMPUTER READABLE FORM:
 13 MEDIUM TYPE: Floppy disk
 14 COMPUTER: IBM PC Compatible
 15 OPERATING SYSTEM: PC-DOS/MS-DOS
 16 SOFTWARE: PatentIn Release #1.0, Version #1.50
 17 CURRENT APPLICATION DATA:
 18 APPLICATION NUMBER: US/09/553,222
 19 FILING DATE:
 20 CLASSIFICATION:
 21 PRIOR APPLICATION DATA:
 22 APPLICATION NUMBER: US/08/646,936
 23 FILING DATE:
 24 ATTORNEY/AGENT INFORMATION:
 25 NAME: Weber, Kenneth A.
 26 REGISTRATION NUMBER: 31,679
 27 REFERENCE/DOCKET NUMBER: 015290-24,000
 28 TELECOMMUNICATION INFORMATION:
 29 TELEPHONE: (415) 576-0297
 30 TELEFAX: (415) 576-0400
 31 INFORMATION FOR SEQ ID NO. 1:
 32 SEQUENCE CHARACTERISTICS:
 33 LENGTH: 239 amino acids
 34 TYPE: amino acid
 35 TOPOLOGY: linear
 36 MOLECULE TYPE: protein
 37 US-09-503-222-2

Query Match 99.6% Score 1270 CB 12 10000 2301
 Best Local Similarity 200.0% Pseq. No. 350 120 148 99
 Matches 237: Conservative 0: Miscellaneous 0: Tolerant 0: 148 99
 QY 2 SKDEELFTAVVPTAVPTKHWVNHKFSVSDGRTATVYKLTKEIYTKGLVWVWPTVY 41
 DE 2 SKDEELFTAVVPTAVPTKHWVNHKFSVSDGRTATVYKLTKEIYTKGLVWVWPTVY 41
 QY 62 TTSTSVGQFSRYTQSKRHHDEPKSKAMPDQYQESLTFEEQDNKTRAPVYQDTLNS 121
 DE 63 TTSTSVGQFSRYTQSKRHHDEPKSKAMPDQYQESLTFEEQDNKTRAPVYQDTLNS 121
 QY 122 RIELKQIDKEKDNITLCHKLEYNKSNHNYDADKQKQIKRNEKIKPKIKEDSVQERDH 181
 DE 123 RIELKQIDKEKDNITLCHKLEYNKSNHNYDADKQKQIKRNEKIKPKIKEDSVQERDH 181
 QY 182 YOONTFGRGPVLPDPNHVLTQSALSQDKPKRHHVLLSEVTAAGTTHOMDEYK 236
 DE 183 YOONTFGRGPVLPDPNHVLTQSALSQDKPKRHHVLLSEVTAAGTTHOMDEYK 236

Search completed: July 16, 2003, 16:15:28
 Job time : 16 secs

Sequence version: 1.000
Copyright (c) 1993 - 2003, GenCorp Inc.

AM protein - protein search, using sw model

Run On: July 16, 2003, 16:00:47 : Search time 35 seconds
Goldblatt alignment
964,742 Million cell updates/sec

Title: US-09-967-301-4
Perfect score: 1276
Sequence: 1 MSKSELFQGVVILVELDG.....VLKFFVTAAGTSCMEILYK 248

Scoring table: SLSCM62
Gapop 10.0 , Gapext 0.5

Searched: 905470 seqs, 124250620 residues

Total number of hits satisfying chosen parameters: 90-470

Minimum PB seq length: 0
Maximum PB seq length: 2500000000

Post-processing: Minimum Match 90
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1992.DAT*
2: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1991.DAT*
3: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1990.DAT*
4: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1989.DAT*
5: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1988.DAT*
6: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1987.DAT*
7: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1986.DAT*
8: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1985.DAT*
9: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1984.DAT*
10: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1983.DAT*
11: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1982.DAT*
12: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1981.DAT*
13: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1980.DAT*
14: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1979.DAT*
15: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1978.DAT*
16: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1977.DAT*
17: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1976.DAT*
18: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1975.DAT*
19: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1974.DAT*
20: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1973.DAT*
21: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1972.DAT*
22: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1971.DAT*
23: /SIDS2/seqdata/geneseq/geneseq-emb1/AA1970.DAT*

Pred. No. is the number of results predicted by BLAST. No. of hits
score greater than or equal to the score of the best hit being returned
and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Match	Length	DB ID	Description
1	1270	99.5	238	19	AAW76106 A. victoria green
2	1266	99.2	238	19	AAW22101 Acetone victoria
3	1266	99.2	238	19	AAW5076 A. victoria green
4	1266	99.2	238	20	AAW96328 Acetone victoria
5	1266	99.2	238	23	AAW16042 Acetone victoria
6	1266	99.2	247	20	AAW96329 Acetone victoria
7	1266	99.2	501	18	AAW96329 GFP variants S65T
8	1262	98.9	236	17	AAW5304 Green fluorescent
9	1262	98.9	236	18	AAW22097 Acetone victoria
10	1262	98.9	236	18	AAW24232 Acetone victoria

11	1262	98.9	236	19	AAW76371 A. victoria green
12	1262	98.9	236	19	AAW76106 A. victoria green
13	1262	98.9	236	19	AAW40479 Wild-type green fl
14	1262	98.9	236	22	AAW73552 Acetone victoria
15	1262	98.9	236	23	AAW16038 GFP fusion protein
16	1262	98.9	432	23	AAW96320 GFP fusion protein
17	1262	98.9	431	23	AAW96321 GFP fusion protein
18	1262	98.9	430	23	AAW96322 GFP fusion protein
19	1262	98.9	430	23	AAW96323 GFP fusion protein
20	1262	98.9	430	23	AAW96324 GFP fusion protein
21	1262	98.9	554	19	AAW48661 GR fusion protein
22	1262	98.9	553	19	AAW48662 GR fusion protein
23	1262	98.9	675	23	AAW76672 protein related to
24	1262	98.9	926	23	AAW96335 GFP fusion protein
25	1262	98.9	1070	18	AAW17789 Green fluorescent
26	1261	98.9	238	17	AAW95312 Green fluorescent
27	1261	98.9	238	19	AAW76372 A. victoria green
28	1261	98.9	238	19	AAW76110 A. victoria green
29	1261	98.9	238	19	AAW55079 A. victoria green
30	1261	98.9	238	19	AAW55084 A. victoria green
31	1261	98.9	238	23	AAW16043 Acetone victoria
32	1261	98.9	238	23	AAW16046 Acetone victoria
33	1260	98.7	238	19	AAW55080 A. victoria green
34	1260	98.7	238	23	AAW16044 Acetone victoria
35	1260	98.7	238	23	AAW16045 Acetone victoria
36	1260	98.7	238	23	AAW16047 Acetone victoria
37	1260	98.7	501	18	AAW31877 GFP variants S65C
38	1260	98.7	501	18	AAW31878 GFP variants S65C
39	1259	98.7	514	18	AAW31876 GFP variants S65C
40	1259	98.7	238	19	AAW76109 A. victoria green
41	1259	98.7	238	19	AAW76111 A. victoria green
42	1259	98.7	238	19	AAW76113 A. victoria green
43	1259	98.7	238	19	AAW61950 Green fluorescent
44	1259	98.7	236	21	AAW33876 Synthetic green fl
45	1259	98.7	236	23	AAW16041 Acetone victoria

ALIGNMENTS

RESULT 1					
AAW76106	AAW76106 standard; protein: 238 AA.				
XX	AAW76106				
XX	AAW76106				
XX	18 NOV-1998 (first entry)				
XX	A. victoria green fluorescent protein mutant F64L/S65T.				
XX	Green fluorescent protein:66 GFP; mutant: jol-yllis; excitation: chlorophore.				
XX	Acquired victoria.				
XX	Synthetic.				
XX	Key	Location/Qualifiers			
XX	Misc-difference 64	/label= F64L			
XX	Misc-difference 65	/note= "Wild-type Phe is replaced with Leu"			
XX	Misc-difference 65	/label= S65T			
XX		/note= "Wild-type Ser is replaced by Thr"			
XX	US580487-A.				
XX	09-SEP-1998.				
XX	31-JAN-1997.	97US-0791342.			
XX	01-FEB-1996.	96US-0010950.			
XX	31-JAN-1997.	97US-0791312.			

181 HQQNTPIGDSFVLLPUNHYLSTQSSLSKDPNEKHKKHVKVLLSEVTAAGLHGHMELLYK 238
 AAE16042
 ID AAE16042 standard; Protein: 238 AA.
 AC AAE16042;
 DT 26-MAR-2002 (first entry)
 DE Aequorea victoria GFP mutant protein (SGT)
 KW Green fluorescent protein; cell lineage tracer; protein localization
 KW GFP; fusion tag; gene expression marker; fluorescent energy acceptor;
 KW immunosay; hybridisation assay; fluorescent energy donor; biosensor;
 KW FRET; fluorescence resonance energy transfer; mutant; enzyme.
 XX Aequorea victoria.
 CS Synthetic.
 XX Key: Localization/Qualitative
 XX Misc-Reference 55
 FT /Note: "Wild type Ser substituted with Thr"
 XX W0202190147-A2
 XX 29-NOV-2001.
 XX 17-MAY-2001: 2001WO 051649
 XX 19-MAY-2000: 2000US-0575947.
 XX (CYBR-) UNIV OREGON STATE.
 XX Wachter P, Remington SJ;
 XX WPI: 2002-063084/11.
 XX Now long wavelength engineered fluorescent proteins, useful as markers
 XX for gene expression, tracers of cell lineage or as fusion tags to
 XX monitor protein localization, or in detection assays, and immunoassays
 XX or hybridization assays
 XX Claim 2: Page 1: 161pp; English.
 CC The patent discloses functional engineered fluorescent protein and its
 CC corresponding polynucleotide. The amino acid sequence of the engineered
 CC protein is identical to Aequorea green fluorescent protein (GFP). The
 CC engineered fluorescent proteins of the invention have varied fluorescent
 CC properties and have the ability to respond to ion concentrations via a
 CC change in fluorescent characteristics. They are useful as markers for
 CC gene expression, tracers of cell lineage or as fusion tags to monitor
 CC protein localization within living cells. The engineered fluorescent
 CC proteins are particularly useful for coupling engineered fluorescent
 CC proteins to antibodies, nucleic acids or other receptors for use in
 CC detection assays, e.g. immunoassays or hybridization assays. They are
 CC useful for tracking the movement of proteins in cells or in systems
 CC for detecting induction of transcription and for the simultaneous
 CC measurement of two or more processes within cells. Proteins of the
 CC invention are also useful as fluorescent energy donors or acceptors
 CC as well as biosensors for detecting amino acids. They are also useful in
 CC fluorescence resonance energy transfer (FRET). The crystal structure
 CC of the GFP is useful for designing mutants having altered fluorescent
 CC characteristics which are particularly useful to identify amino acids
 CC whose substitution alters fluorescent properties of the protein. The
 CC crystal structure of the GFP is also useful for designing mutants having
 CC altered amino binding characteristics which are particularly useful for
 CC identifying amino acids whose substitution alters the specificity and
 CC affinity of the binding site to various amino acids, and for monitoring amino
 CC binding and therefore the concentration of the amino acid. The present
 CC sequence is Aequorea victoria GFP mutant protein (SGT).
 CC Note: This sequence is not shown in the specification but is derived

CC from Aequorea victoria wild type green fluorescent protein shown in
 CC Figure 3 of the specification (AAE16039).
 XX Sequence 238 AA:
 Query Match 99.2% Score 1266; DB 23; Length 238;
 Best Local Similarity 99.2%; Pos. No. 1-30-124;
 Matches 236; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVPTLAVELLDGVNNGHKESVSGSGHGSDATYCKLIKFKICTGKLVWPTL 60
 DE 1 MSKGEELFTGVPTLAVELLDGVNNGHKESVSGSGHGSDATYCKLIKFKICTGKLVWPTL 60
 QY 61 VTITTYGVQCFSEYVDHKKHDFKSAKPEGYVCEFTIFFKDDGNYKTRAEVAFEGTIV 120
 DE 61 VTITTYGVQCFSEYVDHKKHDFKSAKPEGYVCEFTIFFKDDGNYKTRAEVAFEGTIV 120
 QY 121 NRIELKGIDKEKGNLCHHKLEYNYSNHYVTMDKQKNGIKYFKIRHNTENGVQLAD 180
 DE 121 NRIELKGIDKEKGNLCHHKLEYNYSNHYVTMDKQKNGIKYFKIRHNTENGVQLAD 180
 QY 181 HQQNTPIGDSFVLLPUNHYLSTQSSLSKDPNEKHKKHVKVLLSEVTAAGLHGHMELLYK 238
 DE 181 HQQNTPIGDSFVLLPUNHYLSTQSSLSKDPNEKHKKHVKVLLSEVTAAGLHGHMELLYK 238
 RESULT 5
 AAR9632% Standard: Protein: 237 AA.
 XX AAR9632%
 XX AAR9632%
 XX 26-JUN-1999 (first entry)
 XX Humanised green fluorescent protein
 XX Green fluorescent protein; gfp; jellyfish; Aequorea victoria;
 KW humanisation; reporter gene; substrate; cofactor; beta galactosidase;
 KW firefly luciferase; alkaline phosphatase;
 KW chloramphenicol acetyltransferase; CAT; beta glucuronidase; GUS.
 XX Synthetic.
 CS Aequorea victoria.
 XX W030303997-A1.
 XX 28-JAN-1999.
 XX 16-JUL-1998; 98WO-US14692.
 XX 16-JUL-1997; 97US-083327.
 XX (UNFL) UNIV FLORIDA.
 XX Hauswirth W, Huzyczka N, Zolotukhin S;
 WPI: 1999-132241/11.
 DE N-PSDB: AAX08455.
 XX Humanised green fluorescent protein - used to measure gene
 XX expression and identify transformed cells
 XX Claim 5: Page 142: 152pp; English.
 CC Humanised green fluorescent protein (gfp) genes can be used to
 CC identify transformed cells, to measure gene expression in vitro and
 CC in vivo, to label specific cells in multicellular organisms (e.g. to
 CC study cell lineage's), to label and locate fusion proteins, and to
 CC study intracellular trafficking. Commonly used reporter genes include
 CC beta-galactosidase, firefly luciferase, alkaline phosphatase;
 CC chloramphenicol acetyltransferase (CAT), and beta glucuronidase;
 CC (GUS). However, these have limitations in their use. Frequently,
 CC these reporter genes require the addition of a substrate and the

CC size of certain proteins means that the expression of reporter
 CC fusion proteins can be difficult. The high stimulated GFP
 CC fluorescence is species independent and does not require any
 CC cofactors substrates or additional gene products from Aquorea
 CC victoria as the GFP genes have been homologous, they are
 CC expressed at sufficient levels to be detectable in human cells
 CC unlike previous GFP proteins.

XX SQ Sequence 247 AA:
 Query Match 99.24; Score 1256; DR 18; Length 501
 Best Local Similarity 99.24; Prod. No. 3.9e-124;
 Matches 246; Conservative 6; Mismatches 2; Gaps 0

QY 1 MSKGEELFTGVVILVRLDGVNCHKFSVSGEGDGYGKILKPKICIGTKLFPWPTL 60
 DB 10 MSKGEELFTGVVPLVLDGVNCHKFSVSGEGDGYGKILKPKICIGTKLFPWPTL 60
 QY 51 VTLLTYGVQCFSRYPDHHKHDFPKSAMPEGYVOERTIFFKDGNYKTRAEVKEGTLV 120
 DB 7C VTLLTYGVQCFSRYPDHHKHDFPKSAMPEGYVOERTIFFKDGNYKTRAEVKEGTLV 120
 QY 121 NRLELKGIDPKDGNILGHKLEYNHSHVYIMADQKNGIKVNEKTPKNEZESQVLAD 180
 DB 130 NRLELKGIDPKDGNILGHKLEYNHSHVYIMADQKNGIKVNEKTPKNEZESQVLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSAISKDFNEKPKDMVLEFVTAAGTTHGMDELYK 248
 DB 190 HYQONTPIGDGPVLLPDNHYLSTQSAISKDFNEKPKDMVLEFVTAAGTTHGMDELYK 247

RESULT 7

AAW31879 ID AAW31879 standard; Protein, 501 AA.

AC AAW31879;

XX 03-FEB-1998 (first entry)

XX GFP variants S65T and W7 tandem fluorescent protein construct.
 XX North West Pacific jellyfish; green fluorescent protein, GFP; S65T
 XX fluorescent resonance energy transfer; FRET; enzymatic assay; W7
 XX enzyme inhibitor; enzyme promoter; PCR primers; protease cleavage sites
 XX tandem fluorescent protein construct; blue fluorescent protein.

OS Synthetic.

OS Aquorea victoria.

FA Key Definition/Qualifiers

FA Protein 1..248

FT FT Label= S65T_GFP_variant

FT Misc-difference 65

FT FT Label= S65T

FT FT /note= "wild type Ser substituted with Thr"

FT Peptide 239..264

FT FT Label= linker_moiety

FT FT Label= trypsin_cleavage_site

FT FT Label= calpain_cleavage_site

FT FT Label= enterokinase_cleavage_site

FT FT Label= W7_SFP_variant

FT FT Label= Y66W

FT FT /note= "wild type Tyr substituted with Trp"

FT Misc-difference 409

FT FT Label= N146L

FT FT /note= "wild type Asn substituted with Leu"

FT Misc-difference 416

FT FT Label= M153T

FT FT /note= "wild type Met substituted with Thr"
 FT Misc-difference 425
 FT FT Label= V153A
 FT FT /note= "wild type Val substituted with Ala"
 FT Misc-difference 475
 FT FT Label= N121K
 FT FT /note= "wild type Asn substituted with Lys"
 XX XX
 XX W09728241-A1.
 XX
 XX 07-AUG-1997.
 XX
 XX 31-JAN-1997. 97WC-USC1457.
 XX
 XX 31-JAN-1996. 96US-0594575.
 XX
 XX (AURO-) AUROBA BIOSCIENCES CORP.
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Cubitt A. Helm S. Tsien RY:
 XX WPI: 1997-402615/37.
 XX
 XX tandem fluorescent protein constructs - have donor and acceptor
 XX moieties exhibiting fluorescent linked via cleavable peptide linker,
 XX useful in enzymatic assays

PS Claim 3; Page : 88pp; English.

CC This protein sequence is that of a novel tandem fluorescent protein
 CC construct, made using Aquorea victoria (North West Pacific jellyfish)
 CC green fluorescent protein (GFP) variants S65T and W7. W7 fluoresces
 CC at a shorter wavelength than GFP. The construct comprises a donor
 CC (e.g. S65T) and an acceptor (e.g. W7) fluorescent protein moiety
 CC (donors and acceptors can be green or blue fluorescent proteins), and a
 CC linker coupling them. Preferably, the donor is positioned at the
 CC N-terminus of the polypeptide relative to the acceptor. The linker moiety
 CC is a peptide 5-50 amino acids in length containing a protease cleavage
 CC site. In this example, the linker moiety contains many recognition sites
 CC for proteases, including trypsin, calpain and enterokinase. The donor and
 CC acceptor moieties exhibit fluorescent resonance energy transfer (FRET)
 CC when they are cleaved. The constructs are used in enzymatic assays and
 CC can be used to isolate new enzymes or enzyme inhibitors or promoters. The
 CC specific activity of enzyme (in vivo and in vitro) and compounds altering
 CC enzyme activity can be obtained. FRET and hence activity of specific
 CC compounds is measured from the acceptor or donor moiety or maybe obtained
 CC using a ratio between the two.

CC Note: The present sequence does not appear in the specification; it
 CC has been made by modifying the native GFP sequence, and adding the
 CC linker moiety in the appropriate place.

XX SQ Sequence 501 AA:

Query Match 99.24; Score 1256; DR 18; Length 501
 Best Local Similarity 99.24; Prod. No. 3.9e-124;
 Matches 246; Conservative 6; Mismatches 2; Gaps 0

QY 1 MSKGEELFTGVVILVRLDGVNCHKFSVSGEGDGYGKILKPKICIGTKLFPWPTL 60
 DB 1 MSKGEELFTGVVPLVLDGVNCHKFSVSGEGDGYGKILKPKICIGTKLFPWPTL 60
 QY 61 VTLLTYGVQCFSRYPDHHKHDFPKSAMPEGYVOERTIFFKDGNYKTRAEVKEGTLV 120
 DB 61 VTLLTYGVQCFSRYPDHHKHDFPKSAMPEGYVOERTIFFKDGNYKTRAEVKEGTLV 120
 QY 121 NRLELKGIDPKDGNILGHKLEYNHSHVYIMADQKNGIKVNEKTPKNEZESQVLAD 180
 DB 121 NRLELKGIDPKDGNILGHKLEYNHSHVYIMADQKNGIKVNEKTPKNEZESQVLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSAISKDFNEKPKDMVLEFVTAAGTTHGMDELYK 248
 DB 161 HYQONTPIGDGPVLLPDNHYLSTQSAISKDFNEKPKDMVLEFVTAAGTTHGMDELYK 238

RESULT 8
 AAK05304 standard; Protein: 235 AA.
 XX
 AC AAK05304:
 XX
 CD 02-APR-1997 (first entry)
 XX
 DE Green fluorescent protein.
 XX
 KW Green fluorescent protein: GFP; jellyfish; Aequorea victoria; sea pansy;
 KW Renilla reniformis; differential gene expression; protein localization;
 KW gene expression tracking; fluorescence.
 XX
 CS Aequorea victoria.
 XX
 PN W05623810-A1.
 XX
 PD 08-AUG-1996.
 XX
 PF 13-NOV-1995; 95W-0514434.
 XX
 PR 10-NOV-1994; 94US-012315.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 FI Reia R. Tsien EV;
 XX
 CR WPI: 1956-37137/57.
 XX
 CF N-PSDB: AAK3594.
 XX
 PT New modified Aequorea green fluorescent protein (GFPs) - having
 PT amino acid changes to provide prods. which exhibit different
 PI excitation and emission spectra.
 XX
 PS Disclosure: Page 20-21; 6pp; English.
 XX
 CC This sequence represents the wild type Aequorea victoria (jellyfish)
 CC green fluorescent protein (GFP). The fluorescence of this protein is
 CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
 CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
 CC the one absorbance peak seen in the related GFP from the sea pansy
 CC (Renilla reniformis). This protein is modified to provide the GFPs of
 CC the invention. The modifications were created by substituting the CNA
 CC encoding this sequence to site directed mutagenesis using mutagenic PCR
 CC primers, or hydroxylamine treatment. The GFPs of the invention are
 CC modified to lead to the formation of products with markedly different
 CC excitation and emission spectra. Visibly distinct colours, and increased
 CC intensities of emission make these products useful in a wide variety of
 CC contexts, such as tracking of differential gene expression and protein
 CC localisation. The mutations can also be created to modify the encoded
 CC GFP so that it only possesses one absorbance peak.
 XX
 SQ Sequence 235 AA:
 Query Match 98.9%; Score 1262; DB 17; Length 248;
 Best Local Similarity 98.7%; Pred. No. 3.4e-124;
 Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSKGEELFGVGVPIVELDGVNCHKFSVSGEGDANYGKTKIKFICTSKLPVWPTL 60
 DB 1 MSKGEELFGVGVPIVELDGVNCHKFSVSGEGDANYGKTKIKFICTSKLPVWPTL 60
 QY 61 VTTLTYGVQCFSRYPDHMKRPHDFKSAPEGVVQERTIFFKFGDNKTPAEVKFESDLY 120
 DB 61 VTTLTYGVQCFSRYPDHMKRPHDFKSAPEGVVQERTIFFKFGDNKTPAEVKFESDLY 120
 QY 121 NRLEKLGIDFKEDGNLGGKLEYNYNNSHNYINAKSKQKTKVKNKLNEDGVVLAD 180
 DB 121 NRLEKLGIDFKEDGNLGGKLEYNYNNSHNYINAKSKQKTKVKNKLNEDGVVLAD 180
 QY 181 HYQNTPTGDPVILPDRHYLSTQSALEKSKDPREKRDHNVLLFVTAACITGMBELYK 238

DB 1 HYQNTPTGDPVILPDRHYLSTQSALEKSKDPREKRDHNVLLFVTAACITGMBELYK 238
 RESULT 9
 AAK22097 standard; Protein: 233 AA.
 XX
 AC AAK22097:
 XX
 CD 10-MAR-1995 (first entry)
 XX
 DE Aequorea victoria protein GFP10 mutant S65T.
 XX
 KW Green fluorescent protein: GFP; jellyfish; gfp10 gene; reporter;
 KW humanised expression; mutant.
 XX
 CS Aequorea victoria.
 XX
 PN W09726333-A1.
 XX
 PD 24-JUL-1997.
 XX
 PF 17-JAN-1997; 97W-0500755.
 XX
 PR 16-JAN-1996; 96US-0589201.
 XX
 PA (BYPL) UNIV FLORIDA RES FOUND INC.
 XX
 FI Hauswirth W, Murtyczka N, Zelotukhin S;
 XX
 CR WPI: 1997-36537/75.
 XX
 CF N-PSDB: AAK75698.
 XX
 PT Humanised green fluorescent protein gene - optimised to provide high
 PT level expression in mammalian cells, used e.g. to label or identify
 PT cells, to locate proteins, etc.
 XX
 PS Claim 3; Page 1; 158pp; English.
 XX
 CC This sequence represents a mutant form of the green fluorescent protein
 CC GFP10 in which the Ser at position 65 of the wild-type protein has
 CC been replaced by Thr. Green fluorescent proteins (GFP's) could be used
 CC as reporter molecules as they absorb blue light and emit green light
 CC without requiring any cofactors, substrates, or additional gene products
 CC allowing GFP detection in living cells providing meaningful gene
 CC expression is achieved. By providing humanised GFP, adapted for
 CC expression in mammalian and human cells, problems associated with
 CC wild-type jellyfish GFP e.g. variable and low expression levels should be
 CC overcome.
 CC Note: The present sequence does not appear in the specification. It
 CC has been made by modifying the wild-type GFP10 amino acid sequence found
 CC in AAK31255.
 XX
 SQ Sequence 235 AA:
 Query Match 98.9%; Score 1262; DB 18; Length 238;
 Best Local Similarity 98.7%; Pred. No. 3.4e-124;
 Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSKGEELFGVGVPIVELDGVNCHKFSVSGEGDANYGKTKIKFICTSKLPVWPTL 60
 DB 1 MSKGEELFGVGVPIVELDGVNCHKFSVSGEGDANYGKTKIKFICTSKLPVWPTL 60
 QY 61 VTTLTYGVQCFSRYPDHMKRPHDFKSAPEGVVQERTIFFKFGDNKTPAEVKFESDLY 120
 DB 61 VTTLTYGVQCFSRYPDHMKRPHDFKSAPEGVVQERTIFFKFGDNKTPAEVKFESDLY 120

QY 121 NRLEKLGIDFKEDGNILGKLEKLEYNNSHNHYIMADKOKNGKVNFKIRHNIEDGSGVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGKLEKLEYNNSHNHYIMADKOKNGKVNFKIRHNIEDGSGVOLAD 180
 QY 181 HYQONTPTGGDGVLLPDNRYLSTQSLSKDNEKKEKHNVLEFVTAAGITWGMELDYK 238
 DB 181 HYQONTPTGGDGVLLPDNRYLSTQSLSKDNEKKEKHNVLEFVTAAGITWGMELDYK 238

RESULT 12
 AAW76105
 ID AAW76105 standard: protein: 238 AA
 XX
 AC AAW76105;

ET 18-NOV-1998 (first entry)
 XX A. victoria green fluorescent protein.

XX Green fluorescent protein; GFP; mutants: jellyfish; expression;
 KW chromophore.

XX Aequorea victoria.

XX CS5804347-A.

XX 08-SEP-1998.

XX 31-JAN-1997; 97US-0791322.

XX 01-FEB-1996; 96US-0707600.

XX 31-JAN-1997; 97US-0791322.

XX (STRD) UNIV LELAND STANFORD JUNIOR.

XX Cormack RP, Farkow S, Valdivia RH;

XX WPI: 1998-505643/43.

XX DNA encoding mutant green fluorescent protein: proteins - with

XX greater fluorescence intensity than wild-type proteins, useful for

XX studying gene expression and protein localisation

XX Disclosure: Page 1: 15pp; English.

XX This sequence represents the green fluorescent protein (GFP) from

XX Aequorea victoria. This sequence is used in a method to discover GFP's

XX mutated in the chromophore region that fluoresce more brightly than

XX wild-type GFP upon excitation at 488 nm. Those mutants can be used in a

XX method for analyzing a cell containing the GFP. The GFP can also be used

XX to a protein and used to identify the intracellular localisation of a

XX protein of interest. A regulatory element could be operatively connected

XX to a coding portion encoding a mutant GFP and exposed to an environment

XX stimulates the fluorescence signal from the GFP cell; then measures

XX the effect of the stimulus on the regulatory element.

XX Note: This sequence does not appear in the specification but has

XX been retrieved from the EMBL record contained in W5514.

XX Sequence 238 AA;

XX Query Match 98.9%; Score 1262; DB 19; Length 238;

XX Best Local Similarity 98.7%; Pred. No. 3.4e-124;

XX Matches 238; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVETIVELIGVNSHKFSVSRGEGTATVGEKTHAEPTNTIKRLEPWETIL 60

DB 1 MSKGEELFTGVVETIVELIGVNSHKFSVSRGEGTATVGEKTHAEPTNTIKRLEPWETIL 60

QY 61 VTLTGVGVQCFSPYDPMKRRHDFKSMARFGYVQGTIFPKDKNYKTRAAVKKPCEETILV 120

DB 61 VTLTGVGVQCFSPYDPMKRRHDFKSMARFGYVQGTIFPKDKNYKTRAAVKKPCEETILV 120

QY 121 NRLEKLGIDFKEDGNILGKLEKLEYNNSHNHYIMADKOKNGKVNFKIRHNIEDGSGVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGKLEKLEYNNSHNHYIMADKOKNGKVNFKIRHNIEDGSGVOLAD 180
 QY 181 HYQONTPTGGDGVLLPDNRYLSTQSLSKDNEKKEKHNVLEFVTAAGITWGMELDYK 238
 DB 181 HYQONTPTGGDGVLLPDNRYLSTQSLSKDNEKKEKHNVLEFVTAAGITWGMELDYK 238

RESULT 13
 AAW40475
 ID AAW40475 standard: protein: 249 AA
 XX
 AC AAW40475;

ET 15-SEP-1998 (first entry)
 XX A. victoria green fluorescent protein.

XX Green fluorescent protein; protein kinase; phosphate donor;
 KW phosphorylation; drug screening; receptor-ligand binding; signalling;
 KW protein-protein interaction; kinase activation.

XX Aequorea victoria.

XX Key Location/Qualifiers

XX Protein 1..238

XX Note: "Partial sequence"

XX WS966251; AL

XX 22-JAN-1998.

XX 16-JUL-1997; 97WO-0512410.

XX 16-JUL-1996; 96US-0680872.

XX 16-JUL-1996; 96US-0679850.

XX 16-JUL-1996; 96US-0680876.

XX (REGC) UNIV CALIFORNIA.

XX Cubitt AB, Tsien RY;

XX WPI: 1998-110616/10.

XX N-PSDP: AAW1125.

XX Assays for protein kinase and modulators - using a fluorescent

XX protein substrate which exhibits a different fluorescent property in

XX the phosphorylated and un-phosphorylated state

XX Disclosure: Fig 3: 65pp; English.

XX This sequence represents a green fluorescent protein from Aequorea

XX victoria which is used in a method to determine if a sample contains

XX protein kinase (PK) activity. The method involves contacting the sample

XX with a phosphate donor (2D) and a fluorescent protein (FP) substrate for

XX a PK, the protein substrate comprising a FP moiety and a phosphorylation

XX site for a PK, where the protein substrate exhibits a different

XX fluorescent property in the phosphorylated state than in the

XX un-phosphorylated state. The protein substrate is then excited and the

XX amount of a fluorescent property that differs in the un-phosphorylated

XX state and phosphorylated state is measured, whereby an amount that is

XX consistent with the presence of the protein substrate in its

XX phosphorylated state indicates the presence of PK activity. The method

XX and products can be used in drug screening. They can be used for

XX screening for compounds which affect cellular events, including

XX receptor-ligand binding, protein-protein interactions or kinase

XX activation, which signal to the target kinase.

XX Sequence 238 AA;

XX Query Match 98.9%; Score 1262; DB 19; Length 238;

XX Best Local Similarity 98.7%; Pred. No. 3.4e-124;

Matches 2357 Conservative 11 Mismatches 41 Indels 90 Gaps 30

QY 1 MSKGEELFTGVVPLVLELDGVNCHKFSVSGEGGATYKGLTDRPFCITGKLPVWPTL 60
 DQ 1 MSKGEELFTGVVPLVLELDGVNCHKFSVSGEGGATYKGLTDRPFCITGKLPVWPTL 60
 QY 61 VTITVGVQCFSRYPDHMKRHDFKSAHPGVVQERTIEFKDGNKYKTRAEVKEPNTIV 120
 DQ 61 VTITVGVQCFSRYPDHMKRHDFKSAHPGVVQERTIEFKDGNKYKTRAEVKEPNTIV 120
 QY 121 NRTELKGLDFKEDGNLTGSKLEYNKSHNVYINADKQKNGIKVNEKIRINIEPAGVGLAD 180
 DQ 121 NRTELKGLDFKEDGNLTGSKLEYNKSHNVYINADKQKNGIKVNEKIRINIEPAGVGLAD 180
 QY 181 HYCONTPLGCGPVLLPNNHYLSTQSALSKDPNEKRDMHVLFPVTAAGTHGMEELYK 238
 DQ 181 HYCONTPLGCGPVLLPNNHYLSTQSALSKDPNEKRDMHVLFPVTAAGTHGMEELYK 238

RESULT 14
 AAB73552
 ID AAB73552 standard; Protein: 238 AA.
 XX
 AC AAB73552
 DT 07-AUG-2001 (first entry)
 XX
 DE Wild-type green fluorescent protein (wtGFP).
 XX
 KW Phenotype selection: non-selectable; fusion protein: stable expression;
 KW selectable marker: antibiotic resistance gene; Escherichia coli;
 KW green fluorescent protein; GFP; wtGFP; pGFP.
 XX
 OS Aequorea victoria.
 XX
 PN W020012225-A1.
 XX
 PD 26-APR-2001.
 XX
 PY 29-MAR-2000; 2000WO-058477.
 XX
 PK 21-OCT-1999; 90US-0160461.
 PK 22-FEB-2000; 2000US-0510097.
 XX
 PA (PAND-) PANORAMA RES INT.
 XX
 PI Ballint RE.
 XX
 DR WPI: 2001-282162/29.
 DR N-PSDB: AAB20245.
 XX
 XX
 PT Detailing cells expressing mutant proteins comprises of culturing the
 PT cells transformed with a library of mutagenized protein coding sequences
 PT joined to selector protein, which confers growth under selective
 PT conditions.
 XX
 XX
 TS Example 2: Page 7: 52pp; English.
 XX
 XX
 CC The invention relates to methods whereby variants of proteins which do
 CC not confer selectable phenotypes can be selected for stable expression
 CC in host cells (especially Escherichia coli). The methods can be used to
 CC obtain mutants of a desired protein optimised for expression in the host
 CC cells, for obtaining a mutant protein with enhanced stability relative to
 CC the wild-type protein, and for identifying peptides that stabilise an
 CC unstable protein. The methods all involve expressing the protein of
 CC interest as a fusion with a protein that can confer a selectable
 CC phenotype, such as an antibiotic resistance protein. The transformed
 CC host cells are then grown under selective pressure (i.e. presence of
 CC antibiotic). The cells able to grow under such conditions are those
 CC which contain fusion proteins which are optimised for expression or
 CC which are more stable, as this property will also correlate with an
 CC increased amount or rate of synthesis of the selectable marker. The
 CC invention also discloses mutants of green fluorescent protein (GFP).

selected for increased stability using the method of the invention.
 The present sequence represents wild-type GFP (wtGFP) as engineered by
 plasmid pGFP (Genbank U17997), which was used in an exemplification
 of the invention.
 The present sequence is not shown in the specification, but was
 derived from Genbank accession number U17997.
 XX
 SO Sequence 238 AA.

Query Match 28 94; Score 1262; OR 22; Length 238;
 Best Local Similarity 48.7%; Prod. No. 1.4e-124;
 Matches 245; Conservative 11; Mismatches 21; Indels 90; Gaps 30

QY 1 MSKGEELFTGVVPLVLELDGVNCHKFSVSGEGGATYKGLTDRPFCITGKLPVWPTL 60
 DQ 1 MSKGEELFTGVVPLVLELDGVNCHKFSVSGEGGATYKGLTDRPFCITGKLPVWPTL 60
 QY 61 VTITVGVQCFSRYPDHMKRHDFKSAHPGVVQERTIEFKDGNKYKTRAEVKEPNTIV 120
 DQ 61 VTITVGVQCFSRYPDHMKRHDFKSAHPGVVQERTIEFKDGNKYKTRAEVKEPNTIV 120
 QY 121 NRTELKGLDFKEDGNLTGSKLEYNKSHNVYINADKQKNGIKVNEKIRINIEPAGVGLAD 180
 DQ 121 NRTELKGLDFKEDGNLTGSKLEYNKSHNVYINADKQKNGIKVNEKIRINIEPAGVGLAD 180
 QY 181 HYCONTPLGCGPVLLPNNHYLSTQSALSKDPNEKRDMHVLFPVTAAGTHGMEELYK 238
 DQ 181 HYCONTPLGCGPVLLPNNHYLSTQSALSKDPNEKRDMHVLFPVTAAGTHGMEELYK 238

RESULT 15
 AAB73552
 ID AAB73552 standard; Protein: 238 AA.
 XX
 AC AAB73552
 DT 26-MAR-2002 (first entry)
 XX
 DE Aequorea victoria green fluorescent protein (GFP).
 XX
 KW Green fluorescent protein; cell lineage tracer; protein localisation;
 KW GFP; fusion tag; gene expression marker; fluorescent energy acceptor;
 KW immunosay; hybridisation assay; fluorescent energy donor; biosensor;
 KW FRET; fluorescence resonance energy transfer.
 XX
 OS Aequorea victoria.
 XX
 PN W0200190147-A2.
 XX
 PD 29-NOV-2001.
 XX
 PY 17-MAY-2001; 2001WO-0516149.
 XX
 PK 19-MAY-2000; 2000US-057547.
 XX
 PA (OYOR-) UNIV OREGON STATE.
 XX
 PI Wachter R, Remington SJ.
 XX
 DR WPI: 2002-053084/11.
 DR N-PSDB: AAB26207.
 XX
 XX
 TS New long wavelength engineered fluorescent proteins, useful as markers
 TS for gene expression, tracers of cell lineage or as fusion tags to
 TS monitor protein localization, or in detection assays, e.g. immunoassays
 TS or hybridization assays.
 XX
 TS Claim 27: Fig 3: 18pp; English.
 XX
 CC The patent discloses functional engineered fluorescent protein and its
 CC corresponding polynucleotide. The amino acid sequence of the engineered
 CC protein is identical to Aequorea green fluorescent protein (GFP). The
 CC engineered fluorescent proteins of the invention have varied fluorescent

CC properties and have the ability to respond to ion concentrations via a
CC change in fluorescent characteristics. They are useful as markers for
CC gene expression, tracers of cell lineage or as fusion tags to monitor
CC protein localisation within living cells. The engineered fluorescent
CC proteins are particularly useful for coupling engineered fluorescent
CC proteins to antibodies, nucleic acids or other receptors for use in
CC detection assays, e.g. immunoassays or hybridisation assays. They are
CC useful for tracking the movement of proteins in cells or in systems
CC for detecting induction of transcription and for the simultaneous
CC measurement of two or more processes within cells. Proteins of the
CC invention are also useful as fluorescent energy donors or acceptors
CC as well as biosensors for detecting analytes. They are also useful in
CC fluorescence resonance energy transfer (FRET) in the crystal structure
CC of the GFP is useful for designing mutants having altered fluorescent
CC characteristics which are particularly useful to identify amino acids
CC whose substitution alters fluorescent properties of the protein. The
CC crystal structure of the GFP is also useful for designing mutants having
CC altered amino binding characteristics which are particularly useful for
CC identifying amino acids whose substitution alters the specificity and
CC affinity of the binding site to various analytes and for monitoring amino
CC binding and therefore the concentration of the analyte in the presence
CC sequence is Acquired via a green fluorescent protein.
XX
SQ Sequence 238 AA:
Query Match 38.9% Score 12621 IP 24 Length 238
Best Local Similarity 38.7% E-Exp. No. 3.4e-144
Matches 255; Conservative 11 Mismatches 31 Indels 31 Gaps 11
QY 1 MSKGELEATGVVPEILVLEGGVNHKHSVSFENEDDAVYKGLKKEFCEIGKRGVIMVLECE
DB 1 MSKGELEATGVVPEILVLEGGVNHKHSVSFENEDDAVYKGLKKEFCEIGKRGVIMVLECE
QY 61 VTLLVGVGCTSRYPDHKKRHEKCKSAPFQGVVPEILVLEGGVNHKHSVSFENEDDAVYKGLKKEFCEIGKRGVIMVLECE
DB 61 VTLLVGVGCTSRYPDHKKRHEKCKSAPFQGVVPEILVLEGGVNHKHSVSFENEDDAVYKGLKKEFCEIGKRGVIMVLECE
QY 121 NRLEKLGIDKEDKDNITLCKLEYNNHNVYIMAKKKNVLEKVNFKPENIEFGNVDAL 189
DB 121 NRLEKLGIDKEDKDNITLCKLEYNNHNVYIMAKKKNVLEKVNFKPENIEFGNVDAL 189
QY 131 HYQGNIPFGGGVILPEKNEYSTQSAISKQDNKPKRDEEVLEFVTAAGTTHGMDLKK 238
DB 131 HYQGNIPFGGGVILPEKNEYSTQSAISKQDNKPKRDEEVLEFVTAAGTTHGMDLKK 238

Search completed: July 16, 2003, 16:10:35
Job time : 46 secs

4/23/01

Sequence version 1.1
Copyright (c) 1994-2003. All rights reserved.

CM protein - protein search using sw model

Run on: July 16, 2003, 13:00:47 : Search time 17 seconds
(without optimization)
294,734 Million cell updates/1000

Title: US-09-967-301-2

Perfect score: 1275

Sequence: 1 MSKGEELFRVPHVLELDG.....VLEEVTAAGTTHGMDEYK 436

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250420 residues

Total number of hits satisfying chosen parameters: 88470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 04

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*

- 1: /SID52/qcdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 2: /SID52/qcdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID52/qcdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID52/qcdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID52/qcdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID52/qcdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID52/qcdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID52/qcdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID52/qcdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID52/qcdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID52/qcdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID52/qcdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID52/qcdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID52/qcdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID52/qcdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID52/qcdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID52/qcdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID52/qcdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID52/qcdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID52/qcdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID52/qcdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID52/qcdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID52/qcdata/geneseq/geneseq-emb1/AA2002.DAT.*

pred. No. is the number of results predicted by genome to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1275	100.0	238	17	AAW05304 Green fluorescent
2	1275	100.0	238	18	AAW024232 Aequorea victoria
3	1275	100.0	238	19	AAW767471 A. victoria green
4	1275	100.0	238	19	AAW767195 A. victoria green
5	1275	100.0	238	19	AAW40479 A. victoria green
6	1275	100.0	238	23	AAW73552 Wild-type green f.
7	1275	100.0	238	23	AAW16034 Aequorea victoria
8	1274	99.9	238	19	AAW65084 A. victoria green
9	1272	99.8	238	19	AAW76109 A. victoria green
10	1272	99.8	238	19	AAW76110 A. victoria green

11	1274	99.8	238	19	AAW76111 A. victoria green
12	1272	99.8	238	19	AAW50078 A. victoria green
13	1272	99.8	238	19	AAW50079 A. victoria green
14	1272	99.8	238	23	AAW16042 Aequorea victoria
15	1272	99.8	238	23	AAW16043 Aequorea victoria
16	1272	99.8	501	18	AAW18799 GFP variants 8457
17	1271	99.7	238	17	AAW5312 Green fluorescent
18	1271	99.7	238	18	AAW36057 Green fluorescent
19	1271	99.7	238	18	AAW1295 Aequorea victoria
20	1271	99.7	238	19	AAW76113 A. victoria green
21	1271	99.7	238	19	AAW50081 A. victoria green
22	1271	99.7	238	19	AAW50082 A. victoria green
23	1271	99.7	238	20	AAW43243 Green fluorescent
24	1271	99.7	238	20	AAW88484 Green fluorescent
25	1271	99.7	238	21	AAW31616 Aequorea victoria
26	1271	99.7	238	21	AAW03411 Aequorea victoria
27	1271	99.7	238	21	AAW15661 Humanised GFP, HC
28	1271	99.7	238	21	AAW01602 Aequorea victoria
29	1271	99.7	238	21	AAW99804 Aequorea victoria
30	1271	99.7	238	21	AAW90835 Aequorea victoria
31	1271	99.7	238	21	AAW99723 Aequorea victoria
32	1271	99.7	238	21	AAW99732 Aequorea victoria
33	1271	99.7	238	21	AAW70365 Humanised green f1
34	1271	99.7	238	22	AAW73554 Green fluorescent
35	1271	99.7	238	22	AAW85699 A. victoria green
36	1271	99.7	238	23	AAW66199 A. victoria green
37	1271	99.7	238	23	AAW14600 Aequorea victoria
38	1271	99.7	238	23	AAW16041 Aequorea victoria
39	1271	99.7	238	23	AAW16056 Aequorea victoria
40	1271	99.7	238	21	AAW69283 Amino acid sequenc
41	1270	99.6	238	17	AAW05306 Green fluorescent
42	1270	99.6	238	19	AAW59822 Green fluorescent
43	1270	99.6	238	19	AAW50086 A. victoria green
44	1270	99.6	238	19	AAW50082 A. victoria green
45	1270	99.6	238	22	AAW73556 Green fluorescent

ALIGNMENTS

RESULT 1
AAW05304
TC AAW05304 Standard; Protein: 238 AA.
XX
AC AAW05304:
XX
DT 02-APR-1997 (first entry)
XX
DE Green fluorescent protein.
XX
KW Green fluorescent protein; GFP; jellyfish; Aequorea victoria; sea pansy;
KW Ectothia reniformis; differential gene expression; protein localisation;
KW gene expression; tracking; fluorescence.
XX
CS Aequorea victoria.
XX
FN WC952382.0-AL.
XX
TC 06-AUG-1996.
XX
PF 15-NOV-1995; 95WC-0514642.
XX
PR 10-NOV-1994; 94CS-0337455.
XX
(RSCC) UNIV CALIFORNIA.
XX
HELM P. Tsien RV:
XX
WPI: 1996-171470/17.
XX
N-PSDB: AAT39644.
XX
New modified Aequorea green fluorescent polypeptide(s) having
PT amino acid changes to provide prods. which exhibit different

21 excitation and emission spectra
XX Disclosure: Page 20-21: 9pp: English.
XX
XX this sequence represents the wild type Aequorea victoria (Jellyfish)
CC green fluorescent protein (GFP). The fluorescence of this protein is
CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
CC the one absorbance peak seen in the related GFP from the sea pansy
CC (Renilla reniformis). This protein is modified to provide the GFPs of
CC the invention. The modifications were created by substituting the CNA
CC encoding this sequence to site directed mutagenesis using mutagenic PCR
CC primers, or hydroxylamine treatment. The GFPs of the invention are
CC modified to lead to the formation of products with markedly different
CC excitation and emission spectra. Visibly distinct colours and increased
CC intensities of emission make these products useful in a wide variety of
CC contexts, such as tracking of differential gene expression and protein
CC localisation. The mutations can also be created to modify the encoded
CC GFP so that it only possesses one absorbance peak.
XX
XX Sequence 238 AA:
XX
XX Query Match: 100.0%; Score 1275; DB 18; Length 238;
XX Best Local Similarity: 100.0%; Pred. No. 4.6e-124;
XX Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MSKGEELFTSVPTLVVPLVDELGDVNGHKEFSVSGEGDATYKTLKELDTTKELVPEWPTL 60
XX 1 MSKGEELFTSVPTLVVPLVDELGDVNGHKEFSVSGEGDATYKTLKELDTTKELVPEWPTL 60
XX 61 VTFSYGVQCFESRYPDINKSHDFKSAPEGYVQERTIFFKDGNGYKTRAEVKEGDIIV 120
XX 61 VTFSYGVQCFESRYPDINKSHDFKSAPEGYVQERTIFFKDGNGYKTRAEVKEGDIIV 120
XX 61 VTFSYGVQCFESRYPDINKSHDFKSAPEGYVQERTIFFKDGNGYKTRAEVKEGDIIV 120
XX 121 NRLEKGDIDKEDGNILGHKLEYNNSHNYIMADKOKNGIKVNFKIRNTHNSVOLAD 180
XX 121 NRLEKGDIDKEDGNILGHKLEYNNSHNYIMADKOKNGIKVNFKIRNTHNSVOLAD 180
XX 181 HYQONTPIGDPVLLPDNHYLSTQSALSKEPNKROHNVLEFVTAAGTTHGMDELYK 238
XX 181 HYQONTPIGDPVLLPDNHYLSTQSALSKEPNKROHNVLEFVTAAGTTHGMDELYK 238
XX
XX RESULT 3
XX AAW76371 standard: Protein: 238 AA.
XX AC AAW76371;
XX AC AAW76371;
XX DT 03-DEC-1996 (first entry)
XX XX A. victoria green fluorescent protein.
XX XX Green fluorescent protein; GFP; Jellyfish; detection: mutant; activity:
XX KW distribution.
XX XX Aequorea victoria.
XX OS Aequorea victoria.
XX PN JP10234382-A.
XX XX 08-SEP-1998.
XX XX 27-FEB-1997; 97CP-0062370.
XX XX 27-FEB-1997; 97JP-0062370.
XX XX (CHINA) DINABEKU KENKYUSHO KK.
XX XX

21 excitation and emission spectra
XX Disclosure: Page 20-21: 9pp: English.
XX
XX this sequence represents the wild type Aequorea victoria (Jellyfish)
CC green fluorescent protein (GFP). The fluorescence of this protein is
CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
CC the one absorbance peak seen in the related GFP from the sea pansy
CC (Renilla reniformis). This protein is modified to provide the GFPs of
CC the invention. The modifications were created by substituting the CNA
CC encoding this sequence to site directed mutagenesis using mutagenic PCR
CC primers, or hydroxylamine treatment. The GFPs of the invention are
CC modified to lead to the formation of products with markedly different
CC excitation and emission spectra. Visibly distinct colours and increased
CC intensities of emission make these products useful in a wide variety of
CC contexts, such as tracking of differential gene expression and protein
CC localisation. The mutations can also be created to modify the encoded
CC GFP so that it only possesses one absorbance peak.
XX
XX Sequence 238 AA:
XX
XX Query Match: 100.0%; Score 1275; DB 18; Length 238;
XX Best Local Similarity: 100.0%; Pred. No. 4.6e-124;
XX Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MSKGEELFTSVPTLVVPLVDELGDVNGHKEFSVSGEGDATYKTLKELDTTKELVPEWPTL 60
XX 1 MSKGEELFTSVPTLVVPLVDELGDVNGHKEFSVSGEGDATYKTLKELDTTKELVPEWPTL 60
XX 61 VTFSYGVQCFESRYPDINKSHDFKSAPEGYVQERTIFFKDGNGYKTRAEVKEGDIIV 120
XX 61 VTFSYGVQCFESRYPDINKSHDFKSAPEGYVQERTIFFKDGNGYKTRAEVKEGDIIV 120
XX 61 VTFSYGVQCFESRYPDINKSHDFKSAPEGYVQERTIFFKDGNGYKTRAEVKEGDIIV 120
XX 121 NRLEKGDIDKEDGNILGHKLEYNNSHNYIMADKOKNGIKVNFKIRNTHNSVOLAD 180
XX 121 NRLEKGDIDKEDGNILGHKLEYNNSHNYIMADKOKNGIKVNFKIRNTHNSVOLAD 180
XX 181 HYQONTPIGDPVLLPDNHYLSTQSALSKEPNKROHNVLEFVTAAGTTHGMDELYK 238
XX 181 HYQONTPIGDPVLLPDNHYLSTQSALSKEPNKROHNVLEFVTAAGTTHGMDELYK 238
XX
XX RESULT 2
XX AAW24212 standard: Protein: 238 AA.
XX AC AAW24212;
XX AC AAW24212;
XX DT 02-FEB-1998 (first entry)
XX DE Aequorea victoria green fluorescent protein (GFP).
XX XX North West Pacific Jellyfish; green fluorescent protein; GFP; S60G;
XX KW fluorescent resonance energy transfer; FRET; enzymatic assay;
XX KW enzyme inhibitor; enzyme promoter; Tandem fluorescent protein construct;
XX KW blue fluorescent protein; protease cleavage site.
XX
XX Aequorea victoria.
XX OS Aequorea victoria.
XX PN W09726261-A1.
XX XX 07-AUG-1997.
XX XX 31-JAN-1997; 97WO-US01457.
XX XX 31-JAN-1996; 96US-0594575.
XX XX (AUCO-) AUCORA BIOSCIENCES CORP.
XX PA (REGC) UNIV CALIFORNIA.
XX PI Cubitt A, Heim R, Tsien RY.
XX XX

DR WPI: 1998-065641/46.
 IR X-PSCH: AAW61751.
 XX Green fluorescent protein useful for detecting promoter and
 PT targeting activity in a cell
 XX
 PS Disclosed in: Page 6-7; 11pp; Japanese.
 XX
 CC This sequence represents a wild-type green fluorescent protein, GFP,
 CC isolated from *Aequorea victoria*. This protein is used in the
 CC construction of a mutant protein which emits fluorescence 4 times
 CC higher than the conventional fluorescent proteins. The nucleic acid can
 CC be used for the recombinant production of the fluorescent protein and
 CC also for the determination of activity of a promoter to be tested.
 CC comprising steps of introducing the above vector to a host cell and
 CC detecting the fluorescence emitted from the cell. They can additionally
 CC be used for detecting the targeting activity in the cell of an amino
 CC acid sequence to be tested in which the above fluorescent protein is
 CC introduced to a cell and the distribution of the fluorescent protein in
 CC the cell is observed.
 XX
 SQ Sequence 238 AA:
 Query Match: 100.0%; Score 1275; 98 19; Length 238;
 Best Local Similarity 100.0%; Pied. No. 4.6e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPLVELDGVNKHKFSVSGEGSDATYGLKLFETGKLPVNPPTL 60
 DB 1 MSKGEELFTGVVPLVELDGVNKHKFSVSGEGSDATYGLKLFETGKLPVNPPTL 60
 QY 61 VTFSYGVOCFSYPPDHMKRHDFPKSAMPEGYVQERTIFFKDCGNYKTRAEKFEEDTLV 120
 DB 61 VTFSYGVOCFSYPPDHMKRHDFPKSAMPEGYVQERTIFFKDCGNYKTRAEKFEEDTLV 120
 QY 121 NRLEKKGIDKEGNTLGHKLEYNVNSHNYINADKOKNISKVNFKIPNINIGSVGLAD 180
 DB 121 NRLEKKGIDKEGNTLGHKLEYNVNSHNYINADKOKNISKVNFKIPNINIGSVGLAD 180
 QY 181 HYQONTPIGPGVLLPENHYLSTOSALSKEPNKREHMVLEFVTAAGTTHGMDELYK 238
 DB 181 HYQONTPIGPGVLLPENHYLSTOSALSKEPNKREHMVLEFVTAAGTTHGMDELYK 238
 RESULT 5
 AAW76105
 ID AAW76105 standard; protein: 238 AA.
 XX AC AAW76105;
 XX DT 18-NOV-1998 (first entry)
 XX DE A. victoria green fluorescent protein.
 XX KW Green fluorescent protein: GFP; mutant: jellyfish; excitatory;
 XX KW chromophore.
 XX OS *Aequorea victoria*.
 XX PR US2634387-A.
 XX PC 08-SEP-1998.
 XX PE 31-JAN-1997; 97US-0791332.
 XX PR 01-FEB-1996; 96US-0610950.
 XX PR 31-JAN-1997; 97US-0791332.
 XX PA (STED) UNIV DELAND SCANFORD JUNIOR.
 XX PC Cermack BP, Falkow S, Valdivia RH:
 XX WP: 1998-065641/43.

XX DNA encoding mutant green fluorescent pigment proteins - with
 PT treated fluorescence intensity than wild-type proteins. useful for
 PT studying gene expression and protein localisation
 XX
 PS Disclosed in: Page 6-7; 15pp; English.
 XX
 CC This sequence represents the green fluorescent protein (GFP) from
 CC *Aequorea victoria*. This sequence is used in a method to discover GFP's
 CC mutated in the chromophore region that fluoresce more brightly than
 CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a
 CC method for analyzing a cell containing the GFP. The GFP can also be used
 CC to a protein and used to identify the intracellular localisation of a
 CC protein of interest. A regulatory element could be operatively connected
 CC to a coding portion encoding a mutant GFP and exposed to an environmental
 CC stimulus. The fluorescence signal from the cell then measures
 CC the effect of the stimulus on the regulatory element.
 CC Note: This sequence does not appear in the specification but has
 CC been retrieved from the EMBL record contained in W05304.
 XX
 SQ Sequence 238 AA:
 Query Match: 100.0%; Score 1275; 98 19; Length 238;
 Best Local Similarity 100.0%; Pied. No. 4.6e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPLVELDGVNKHKFSVSGEGSDATYGLKLFETGKLPVNPPTL 60
 DB 1 MSKGEELFTGVVPLVELDGVNKHKFSVSGEGSDATYGLKLFETGKLPVNPPTL 60
 QY 61 VTFSYGVOCFSYPPDHMKRHDFPKSAMPEGYVQERTIFFKDCGNYKTRAEKFEEDTLV 120
 DB 61 VTFSYGVOCFSYPPDHMKRHDFPKSAMPEGYVQERTIFFKDCGNYKTRAEKFEEDTLV 120
 QY 121 NRLEKKGIDKEGNTLGHKLEYNVNSHNYINADKOKNISKVNFKIPNINIGSVGLAD 180
 DB 121 NRLEKKGIDKEGNTLGHKLEYNVNSHNYINADKOKNISKVNFKIPNINIGSVGLAD 180
 QY 181 HYQONTPIGPGVLLPENHYLSTOSALSKEPNKREHMVLEFVTAAGTTHGMDELYK 238
 DB 181 HYQONTPIGPGVLLPENHYLSTOSALSKEPNKREHMVLEFVTAAGTTHGMDELYK 238
 RESULT 5
 AAW40479
 ID AAW40479 standard; protein: 238 AA.
 XX AC AAW40479;
 XX DT 15-SEP-1998 (first entry)
 XX DE A. victoria green fluorescent protein.
 XX KW Green fluorescent protein: protein kinase; phosphate donor;
 XX KW phosphorylation; drug screening; receptor-ligand binding; signalling;
 XX KW protein-protein interaction; kinase activation.
 XX OS *Aequorea victoria*.
 XX PR WO9502571-A1.
 XX PC 22-JAN-1998.
 XX PE 16-JUL-1997; 97WO-US12410.
 XX PR 16-JUL-1996; 96US-0680472.
 XX PR 16-JUL-1996; 96US-0679865.
 XX PR 16-JUL-1996; 96US-0680476.

(REGC) UNIV CALIFORNIA.

XX Qubitt Ab, Tsien KY;

XX WPI: 2993-110416/12.

DR N-PSDB: AAV11125.

XX Assays for protein kinase and modulators using a fluorescent

PT protein substrate which exhibits a different fluorescent property in

PT the phosphorylated and un-phosphorylated state

XX PS Disclosure: Fig 3: 55pp; English.

XX This sequence represents a green fluorescent protein from Aequorea victoria which is used in a method to determine if a sample contains protein kinase (PK) activity. The method involves contacting the sample with a phosphate donor (PD) and a fluorescent protein (FP) substrate for a PK, the protein substrate comprising a fluorophore and a phosphorylation site for a PK, where the protein substrate exhibits a different fluorescent property in the phosphorylated state than in the un-phosphorylated state. The protein substrate is then excited and the amount of a fluorescent property that differs in the un-phosphorylated state and phosphorylated state is measured, whereby an amount that is consistent with the presence of the protein substrate in its un-phosphorylated state indicates the presence of PK activity. The method and products can be used in drug screening. They can be used for screening for compounds which affect cellular events, including receptor-ligand binding, protein-protein interactions or kinase activation, which signal to the target kinase.

XX SQ Sequence 238 AA:

Query Match 100.0%; Score 1275; DB 10; Length 238;

Best Local Similarity 100.0%; Pred. No. 4, 60-124;

Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVILVELDGVNGHFKFSVSGEGGATYGLKILKFKICTGKLVPEWPL 60
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 1 MSKGEELFTGVVILVELDGVNGHFKFSVSGEGGATYGLKILKFKICTGKLVPEWPL 60
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 61 VTFESYGVOCFSRYPDHMKRPHDFKFSAMPEGYVOERTIFFKDGNTKTRAFKFEQDTLV 120
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 61 VTFESYGVOCFSRYPDHMKRPHDFKFSAMPEGYVOERTIFFKDGNTKTRAFKFEQDTLV 120
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 121 NRIELKSGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIFGSVOLAD 180
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 121 NRIELKSGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIFGSVOLAD 180
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 181 HYQONTPIGDGPVLLPNHNYLSQSALSKDPNKRKHNVILLEFVTAGITHGMDELYK 238
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 181 HYQONTPIGDGPVLLPNHNYLSQSALSKDPNKRKHNVILLEFVTAGITHGMDELYK 238
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6

AAB73552

ID AAB73552 standard; Protein: 238 AA.

XX AC AAB73552;

XX 07-AUG-2003; (first entry)

XX DE Wild-type green fluorescent protein (wtGFP).

XX Phenotype selection; non-selectable; fusion protein; stable expression;
 KW selectable marker; antibiotic resistance gene; Escherichia coli;
 KW green fluorescent protein; GFP; wtGFP; pGFP.

CS Aequorea victoria.

XX PN W0200129225-A1.

XX PD 26-APR-2003.

XX

29-MAR-2000; 2000W0-US08477.

XX 21-OCT-1999; 90US-0160461.

XX 22-FEB-2000; 2000US-051097.

XX (PANG-) PANORAMA RES INC.

XX Salint RF:

XX WPI: 2001-292162/29.

XX N-PSDB: AAB20245.

XX Obtaining cells expressing mutant protein, comprises selecting from
 PT cells transformed with library of mutagenized protein coding sequences
 PT fused to selector protein, which confers growth under selective
 PT conditions.

XX Example 2: Page 1: 55pp; English.

XX The invention relates to methods whereby variants of proteins which do
 CC not confer selectable phenotypes can be selected for stable expression
 CC in host cells (especially Escherichia coli). The methods can be used to
 CC obtain mutants of a desired protein optimised for expression in the host
 CC cells, for obtaining a mutant protein with enhanced stability relative to
 CC the wild-type protein, and for identifying peptides that stabilise an
 CC unstable protein. The methods all involve expressing the protein of
 CC interest as a fusion with a protein that can confer a selectable
 CC phenotype, such as an antibiotic resistance protein. The transformed
 CC host cells are then grown under selective pressure (e.g., presence of
 CC antibiotic). The cells able to grow under such conditions are those
 CC which contain fusion proteins which are optimised for expression or
 CC which are more stable, as this property will also correlate with an
 CC increased amount or rate of synthesis of the selectable marker. The
 CC invention also discloses mutants of green fluorescent protein (GFP)
 CC selected for increased stability using the method of the invention.
 CC The present sequence represents wild-type GFP (wtGFP) as encoded by
 CC the plasmid pGFP (Genbank U17597), which was used in an exemplification
 CC of the invention.

XX The present sequence is not shown in the specification, but was

XX derived from GenBank accession number U17597.

XX SQ Sequence 238 AA:

Query Match 100.0%; Score 1275; DB 22; Length 238;

Best Local Similarity 100.0%; Pred. No. 4, 60-124;

Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVILVELDGVNGHFKFSVSGEGGATYGLKILKFKICTGKLVPEWPL 60
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 1 MSKGEELFTGVVILVELDGVNGHFKFSVSGEGGATYGLKILKFKICTGKLVPEWPL 60
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 61 VTFESYGVOCFSRYPDHMKRPHDFKFSAMPEGYVOERTIFFKDGNTKTRAFKFEQDTLV 120
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 61 VTFESYGVOCFSRYPDHMKRPHDFKFSAMPEGYVOERTIFFKDGNTKTRAFKFEQDTLV 120
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 121 NRIELKSGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIFGSVOLAD 180
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 121 NRIELKSGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIFGSVOLAD 180
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 181 HYQONTPIGDGPVLLPNHNYLSQSALSKDPNKRKHNVILLEFVTAGITHGMDELYK 238
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 181 HYQONTPIGDGPVLLPNHNYLSQSALSKDPNKRKHNVILLEFVTAGITHGMDELYK 238
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7

AAB73538

ID AAB73538 standard; Protein: 238 AA.

XX AC AAB73538;

XX 26-MAR-2002; (first entry)

XX DE Aequorea victoria green fluorescent protein (GFP).

CC NOTE: This sequence does not appear in the specification and has
 CC been constructed from the wild-type sequence represented in AA04179.
 XX
 SQ Sequence 238 AA;
 Query Match 99.9%; Score 1274; DB 15; Length 238;
 Rest Local Similarity 99.6%; Pred. No. 9,3e-124;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGSGEDATYGLKLFCTTCKLPVPEPTL 60
 DE 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGSGEDATYGLKLFCTTCKLPVPEPTL 60
 QY 61 VTTSYGVGCFKSRYPDMKKRDEFFKSAMPEGVQERTIFFKDDGNYKTSAEVKEFEDTIV 120
 DE 61 VTTSYGVGCFKSRYPDMKKRDEFFKSAMPEGVQERTIFFKDDGNYKTSAEVKEFEDTIV 120
 QY 121 NRTELKGDIFKEDGNILGHKLEYNNSHNHYIMACKQKNGIKVNEKIRHINIEDGSGVQLAD 180
 DE 121 NRTELKGDIFKEDGNILGHKLEYNNSHNHYIMACKQKNGIKVNEKIRHINIEDGSGVQLAD 180
 QY 181 HYQONTPIGCGPVLLPDNHYLSTQSAISKDPREKRDHMYLLFEVTAAGITHRMDELYK 238
 DE 181 HYQONTPIGCGPVLLPDNHYLSTQSAISKDPREKRDHMYLLFEVTAAGITHRMDELYK 238
 RESULT 9
 AA076109
 ID AA076109 standard: protein; 218 AA.
 XX
 AC AA076109;
 DT 18-NOV-1998 (first entry)
 DE A. victoria green fluorescent protein mutant S26A.
 XX
 KW Green fluorescent protein: GFP; mutant: jellyfish; excitation:
 KW chromophore.
 XX
 CS Aequorea victoria.
 XX Synthetic.
 XX
 FH KEY location/Qualifiers
 FT Misc-difference 72
 FT /label= S26A
 FT /note= "Wild-type Ser is replaced by Ala"
 XX
 PN US58043e7-A.
 XX
 PO 08-SEP-1998.
 XX
 PF 31-JAN-1997; 97US-0791332.
 XX
 PR 01-FEB-1996; 96US-0010960.
 PR 31-JAN-1997; 97US-0791332.
 XX
 PA (STRC) UNIV LELAND STANFORD JUNIOR.
 XX
 PL Cormack BP, Falkow S, Valdivia RH;
 XX
 DR WPI; 1993-505643/43.
 XX
 CC DNA encoding mutant green fluorescent pigment proteins - with
 CC greater fluorescence intensity than wild-type proteins. useful for
 CC studying gene expression and protein localisation.
 XX
 PS Claim 5; Page 1; 15pp; English.
 XX
 CC This sequence represents a mutant green fluorescent protein (GFP) from
 CC Aequorea victoria in which a Ser residue at position 12 in the wild
 CC type protein is replaced by an Ala. These mutations occur in the
 CC chromophore region. GFP mutants are used in a method to discover GFP's
 CC mutated in the chromophore region that fluoresce more brightly than

CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a
 CC method for analyzing a cell containing the GFP. The GFP can also be
 CC fused to a protein and used to identify the intracellular localisation of
 CC a protein of interest. A regulatory element could be operatively
 CC connected to a coding portion encoding a mutant GFP and exposed to an
 CC environmental stimulus. The fluorescence signal from the cell
 CC then measures the effect of the stimulus on the regulatory element.
 CC NOTE: This sequence does not appear in the specification but has been
 CC constructed from the wild-type GFP protein represented in AA076106
 XX
 SQ Sequence 238 AA;
 Query Match 99.8%; Score 1272; DB 15; Length 238;
 Rest Local Similarity 99.6%; Pred. No. 9,3e-124;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGSGEDATYGLKLFCTTCKLPVPEPTL 60
 DE 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGSGEDATYGLKLFCTTCKLPVPEPTL 60
 QY 61 VTTSYGVGCFKSRYPDMKKRDEFFKSAMPEGVQERTIFFKDDGNYKTSAEVKEFEDTIV 120
 DE 61 VTTSYGVGCFKSRYPDMKKRDEFFKSAMPEGVQERTIFFKDDGNYKTSAEVKEFEDTIV 120
 QY 121 NRTELKGDIFKEDGNILGHKLEYNNSHNHYIMACKQKNGIKVNEKIRHINIEDGSGVQLAD 180
 DE 121 NRTELKGDIFKEDGNILGHKLEYNNSHNHYIMACKQKNGIKVNEKIRHINIEDGSGVQLAD 180
 QY 181 HYQONTPIGCGPVLLPDNHYLSTQSAISKDPREKRDHMYLLFEVTAAGITHRMDELYK 238
 DE 181 HYQONTPIGCGPVLLPDNHYLSTQSAISKDPREKRDHMYLLFEVTAAGITHRMDELYK 238
 RESULT 10
 AA076110
 ID AA076110 standard: protein; 218 AA.
 XX
 AC AA076110;
 DT 18-NOV-1998 (first entry)
 DE A. victoria green fluorescent protein mutant S65A.
 XX
 KW Green fluorescent protein: GFP; mutant: jellyfish; excitation:
 KW chromophore.
 XX
 CS Aequorea victoria.
 XX Synthetic.
 XX
 FH KEY location/Qualifiers
 FT Misc-difference 65
 FT /label= S65A
 FT /note= "Wild-type Ser is replaced by Ala"
 XX
 PN US58043e7-A.
 XX
 PO 08-SEP-1998.
 XX
 PF 31-JAN-1997; 97US-0791332.
 XX
 PR 01-FEB-1996; 96US-0010960.
 PR 31-JAN-1997; 97US-0791332.
 XX
 PA (STRC) UNIV LELAND STANFORD JUNIOR.
 XX
 PL Cormack BP, Falkow S, Valdivia RH;
 XX
 DR WPI; 1993-505643/43.
 XX
 CC DNA encoding mutant green fluorescent pigment proteins - with
 CC greater fluorescence intensity than wild-type proteins. useful for
 CC studying gene expression and protein localisation
 XX

PS Claim 6; Page 7; 15pp; English.

XX This sequence represents a mutant green fluorescent protein (GFP) from
XX Aequorea victoria in which a Ser residue at position 65 in the wild
XX type protein is replaced by an Ala. These mutations occur in the
XX chromophore region. GFP mutants are used in a method to discover GFP's
XX mutated in the chromophore region that fluoresce more brightly than
XX wild-type GFP upon excitation at 488 nm. These mutants can be used in a
XX method for analyzing a cell containing the GFP. The GFP can also be
XX fused to a protein and used to identify the intracellular localization of
XX a protein of interest. A regulatory element could be operatively
XX connected to a coding portion encoding a mutant GFP and exposed to an
XX environmental stimulus. The fluorescence signal from the cell
XX then measures the effect of the stimulus on the regulatory element.
XX NOTE: This sequence does not appear in the specification but has been
XX constructed from the wild-type GFP protein represented in AAW76116

XX Sequence 238 AA:

Query Match 99.8%; Score 1272; DB 19; Length 238;
Best Local Similarity 99.6%; Pred. No. 9,36-124;
Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
DY 1 MSKGEELFTGVVPIVLVEELGDNVNGHKFSVSGEGSDATYGLTKLFICTGKLPVNPRL 60
DB 1 MSKGEELFTGVVPIVLVEELGDNVNGHKFSVSGEGSDATYGLTKLFICTGKLPVNPRL 60
DY 61 VTTFSGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKEGCTLV 120
DB 61 VTTFSGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKEGCTLV 120
DY 62 VTTFSGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKEGCTLV 120
DB 62 VTTFSGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKEGCTLV 120
DY 121 NRLEKGIIDFKEDGNLGHKLEYNYNHNYIMADQKNGIKVNFKIRHNEDGSSVQLAD 180
DB 121 NRLEKGIIDFKEDGNLGHKLEYNYNHNYIMADQKNGIKVNFKIRHNEDGSSVQLAD 180
DY 181 HYQONTPIGDSPVLLPDHNYLSTQSALSKDQNEKRDHNVLLFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDSPVLLPDHNYLSTQSALSKDQNEKRDHNVLLFVTAAGITHGMDELYK 238

RESULT 11
AAW76111

ID AAW76111 standard; protein: 238 AA.

XX AAW76111:

XX 18-NOV-1998 (first entry)

XX A. victoria green fluorescent protein mutant: vawc2

XX Green fluorescent protein: GFP; mutant: vallyfish; excitation:
XX chromophore.

XX Aequorea victoria.

XX Synthetic.

XX Key Location/Qualifiers

XX Misc-difference 65

XX /label= V65L

XX /note= "wild-type Val is replaced by Leu"

XX D55824367-A.

XX 08-SEP-1998.

XX 31-JAN-1997; 9705-0791332.

XX 01-FEB-1995; 9605-0010760.

XX 31-JAN-1997; 9705-0791332.

XX (STRD) UNIV LELAND STANFORD JUNIOR.

XX Cormack BP, Faikow S, Valdivia RH;

XX WP1; 1998 505645/43.

XX DNA encoding mutant green fluorescent pigment proteins with
XX altered fluorescence intensity than wild-type proteins, useful for
XX studying gene expression and protein localisation

XX Claim 7; Page 7; 15pp; English.

XX This sequence represents a mutant green fluorescent protein (GFP) from
XX Aequorea victoria in which a Val residue at position 68 in the wild
XX type protein is replaced by a Leu. These mutations occur in the
XX chromophore region. GFP mutants are used in a method to discover GFP's
XX mutated in the chromophore region that fluoresce more brightly than
XX wild-type GFP upon excitation at 488 nm. These mutants can be used in a
XX method for analyzing a cell containing the GFP. The GFP can also be
XX fused to a protein and used to identify the intracellular localisation of
XX a protein of interest. A regulatory element could be operatively
XX connected to a coding portion encoding a mutant GFP and exposed to an
XX environmental stimulus. The fluorescence signal from the cell
XX then measures the effect of the stimulus on the regulatory element.
XX NOTE: This sequence does not appear in the specification but has been
XX constructed from the wild-type GFP protein represented in AAW76106

XX Sequence 238 AA:

Query Match 99.8%; Score 1272; DB 19; Length 238;
Best Local Similarity 99.6%; Pred. No. 9,36-124;
Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
DY 1 MSKGEELFTGVVPIVLVEELGDNVNGHKFSVSGEGSDATYGLTKLFICTGKLPVNPRL 60
DB 1 MSKGEELFTGVVPIVLVEELGDNVNGHKFSVSGEGSDATYGLTKLFICTGKLPVNPRL 60
DY 61 VTTFSGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKEGCTLV 120
DB 61 VTTFSGVQCFSRYPDHMKRHDFPKSAMPEGVYQERTIFFKDDGNYKTRAEVKEGCTLV 120
DY 121 NRLEKGIIDFKEDGNLGHKLEYNYNHNYIMADQKNGIKVNFKIRHNEDGSSVQLAD 180
DB 121 NRLEKGIIDFKEDGNLGHKLEYNYNHNYIMADQKNGIKVNFKIRHNEDGSSVQLAD 180
DY 181 HYQONTPIGDSPVLLPDHNYLSTQSALSKDQNEKRDHNVLLFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDSPVLLPDHNYLSTQSALSKDQNEKRDHNVLLFVTAAGITHGMDELYK 238

RESULT 12

ID AAW5078 standard; protein: 248 AA.

XX AAW5078:

XX 18-SEP-1996 (first entry)

XX A. victoria green fluorescent protein variant S65T.

XX Green fluorescent protein: protein kinase; phosphate donor;
XX phosphorylation: drug screening; receptor-ligand binding; signalling;
XX protein-protein interaction; kinase activation.

XX Aequorea victoria.

XX Synthetic.

XX Key Location/Qualifiers

XX Misc-difference 65

XX /label= S65T

XX /note= "wild type Ser is replaced with Thr"

XX W09802571-A1.

XX 22-JAN-1998.

XX

PF 16-JUL-1997; 97WO-US12410.
 XX 16-JUL-1996; 96US-0680877.
 PR 16-JUL-1996; 96US-0679865.
 XX 16-JUL-1996; 96US-0680876.
 XX (REGC) UNIV CALIFORNIA.
 XX CUBITT AB, Tsien KY;
 XX WPI: 1998-110616/10.
 XX Assays for protein kinase and modulators - using a fluorescent
 PT protein substrate which exhibits a different fluorescent property in
 PT the phosphorylated and un-phosphorylated state
 XX Claim 27: Page -: 65pp; English.
 XX This sequence represents a green fluorescent protein variant from
 CC Aequorea victoria in which a wild-type Ser residue at position 65 is
 CC replaced with a Thr. This variant is used in a method to determine if a
 CC sample contains protein kinase (PK) activity. The method involves
 CC contacting the sample with a phosphate donor (PD) and a fluorescent
 CC protein (FP) substrate for a PK, the protein substrate comprising a FP
 CC moiety and a phosphorylation site for a PK, where the protein substrate
 CC exhibits a different fluorescent property in the phosphorylated state
 CC than in the un-phosphorylated state. The protein substrate is then
 CC excited and the amount of a fluorescent property that differs in the
 CC un-phosphorylated state and phosphorylated state is measured, whereby an
 CC amount that is consistent with the presence of the protein substrate in
 CC its phosphorylated state indicates the presence of PK activity. The
 CC method and products can be used in drug screening. They can be used for
 CC screening for compounds which affect cellular events, including
 CC receptor-ligand binding, protein-protein interactions or kinase
 CC activation, which signal to the target kinase.
 CC NOTE: This sequence does not appear in the specification and has
 CC been constructed from the wild-type sequence represented in AAK40479.
 XX Sequence 238 AA:
 Query Match: 99.8%; Score 1272; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 9,3e-124;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVFLVDELGVNCHKFSVSGSGEDATYCKILKFKICTGKLVFWPTL 60
 DB 1 MSKGEELFTGVVFLVDELGVNCHKFSVSGSGEDATYCKILKFKICTGKLVFWPTL 60
 QY 61 VTTFYGVQCFSPYFDHKKHRLFFKSAMPEGYVQERTIFFKDKGNKYIRAEVKEGTLV 120
 DB 61 VTTFYGVQCFSPYFDHKKHRLFFKSAMPEGYVQERTIFFKDKGNKYIRAEVKEGTLV 120
 QY 121 NRTELKGDIDFKEDSNILGHKLEYNYNHNYIMADKQKNGIKVNFKIRHNEDSSVOLAD 180
 DB 121 NRTELKGDIDFKEDSNILGHKLEYNYNHNYIMADKQKNGIKVNFKIRHNEDSSVOLAD 180
 QY 181 HYQONTPIGDPVLLPDPNHVLTQSALSKDPNFKRHMLLEFVTAAGITGMLEYK 238
 DB 181 HYQONTPIGDPVLLPDPNHVLTQSALSKDPNFKRHMLLEFVTAAGITGMLEYK 238

Protein-protein interaction: kinase activation.
 Aequorea victoria.
 Synthetic.
 Key Location/Organism
 Misc-difference 65 /label= S65A
 /note= "wild type Ser is replaced with Ala"
 WC8902571-A1.
 22-JAN-1998; 97WO-US12410.
 16-JUL-1997; 97WO-US12410.
 16-JUL-1996; 96US-0680877.
 16-JUL-1996; 96US-0679865.
 16-JUL-1996; 96US-0680876.
 (REGC) UNIV CALIFORNIA.
 CUBITT AB, Tsien KY;
 WPI: 1998-110616/10.
 Assays for protein kinase and modulators - using a fluorescent
 protein substrate which exhibits a different fluorescent property in
 the phosphorylated and un-phosphorylated state
 Claim 27: Page -: 65pp; English.
 This sequence represents a green fluorescent protein variant from
 Aequorea victoria in which a wild-type Ser residue at position 65 is
 replaced with a Ala. This variant is used in a method to determine if a
 sample contains protein kinase (PK) activity. The method involves
 contacting the sample with a phosphate donor (PD) and a fluorescent
 protein (FP) substrate for a PK, the protein substrate comprising a FP
 moiety and a phosphorylation site for a PK, where the protein substrate
 exhibits a different fluorescent property in the phosphorylated state
 than in the un-phosphorylated state. The protein substrate is then
 excited and the amount of a fluorescent property that differs in the
 un-phosphorylated state and phosphorylated state is measured, whereby an
 amount that is consistent with the presence of the protein substrate in
 its phosphorylated state indicates the presence of PK activity. The
 method and products can be used in drug screening. They can be used for
 screening for compounds which affect cellular events, including
 receptor-ligand binding, protein-protein interactions or kinase
 activation, which signal to the target kinase.
 NOTE: This sequence does not appear in the specification and has
 been constructed from the wild-type sequence represented in AAK40479.
 XX Sequence 238 AA:
 Query Match: 99.8%; Score 1272; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 9,3e-124;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVFLVDELGVNCHKFSVSGSGEDATYCKILKFKICTGKLVFWPTL 60
 DB 1 MSKGEELFTGVVFLVDELGVNCHKFSVSGSGEDATYCKILKFKICTGKLVFWPTL 60
 QY 61 VTTFYGVQCFSPYFDHKKHRLFFKSAMPEGYVQERTIFFKDKGNKYIRAEVKEGTLV 120
 DB 61 VTTFYGVQCFSPYFDHKKHRLFFKSAMPEGYVQERTIFFKDKGNKYIRAEVKEGTLV 120
 QY 121 NRTELKGDIDFKEDSNILGHKLEYNYNHNYIMADKQKNGIKVNFKIRHNEDSSVOLAD 180
 DB 121 NRTELKGDIDFKEDSNILGHKLEYNYNHNYIMADKQKNGIKVNFKIRHNEDSSVOLAD 180
 QY 181 HYQONTPIGDPVLLPDPNHVLTQSALSKDPNFKRHMLLEFVTAAGITGMLEYK 238
 DB 181 HYQONTPIGDPVLLPDPNHVLTQSALSKDPNFKRHMLLEFVTAAGITGMLEYK 238

CC Figure 3 of the specification (AAE16042).

XX
SQ Sequence 238 AA.

SQTY Ratio 99.84; Score 1272; EB 23; Length 238;
Best Local Similarity 99.64; Pred. No. base-124;
Matches 237; Conservation 2; Mismatches 6; Indels 6; Gaps 0;

QY 1 MSKGEELTGCVVFLIVELDGLVNGHKFSVSSEGSATYSKLTIKFCTITCKIPWPEPTL 60
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
DQ 1 MSKGRELTGVVPVLAVPLGSLVNGCHKFSVSGESGDALYKLTILKEICTTICKLPWPAPTLL 60
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII

QY 61 VTTFSTGYCGRKFEEDHKKRHDFKSAMFEGYVGERTIFFKDDGNKYTRAEVKFESJTLV 120
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
DQ 61 VTTFTYGVCFSFYRCHMKKHDFKSAHFEGYVGERTIFFKDDGNKYTRAEVKFESJTLV 120
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII

QY 121 NATELKGIQFKEDNIGLQKLFYNVNSHRVIMAKOKNGIKVFKRHNEFGSVCAR 180
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
DQ 121 NRTELGIDIFREDGNELOKLEYNVNSHRVIMADKOKNGIKVFKRHNEFGSVIAC 180
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII

QY 183 HYQONTPIGUGPVLLDTNEYLSQSALKSPENKRDHMLVEFFVTAGITHGMDELYK 238
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
DQ 181 HQQNTPTCDGPVLLFDNHYLTSALSNDPENEKRDHMYLLEFVTAGITHGMDELYK 238
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII

RESULT 15
AAE16042
ID AAE16043 standard; Protein: 238 AA.
AC AAE16043;
XX XX
DI 25-MAR-2002 (first entry)
XX XX
DE Aequeorea victoria GFP mutant protein (S65A).
XX XX
KW Green fluorescent proteins: cell lineage tracer; protein localisation;
KW GFP; fusion tag; gene expression marker; fluorescent energy acceptor;
KW Immunassay; hybridisation assay; fluorescent energy donor; biosensor;
KW PRE1; fluorescence resonance energy transfer; mutant; mutagen.
XX OS Aequeorea victoria
OS Synthetic.
XX FH Key Localization/Organifiers
FH FT Misc-difference 65 /note: "Wild type Ser substituted with Ala"
XX XX W020190147-A2.
XX XX 25-NOV-2001.
XX XX 17-MAY-2001; 2001WA-US16-A9.
XX XX 19-MAY-2000; 2000US-0575847.
XX XX {OVR-} UNIV OREGON STATE.
XX XX Wachter R., Remington SJ.
XX XX WPI: 2002-063084/11.
XX XX New long wavelength engineered fluorescent proteins: useful as markers
XX XX for gene expression, tracers of cell lineage or as fusion tags to
XX XX monitor protein localization, or in detection assays, e.g. immunossays
XX XX or hybridization assays .
XX XX Claim 2: Page : 18pp; English.
XX XX The patent discloses functional engineered fluorescent protein and its
XX XX corresponding polynucleotide, the amino acid sequence of the engineered
XX XX protein is identical to Aequeorea green fluorescent protein (GFP). The
XX XX engineered fluorescent proteins of the invention have varied fluorescent
XX XX properties and have the ability to respond to given concentrations via a
XX XX change in fluorescent characteristics, they are useful as markers for
XX XX gene expression, tracers of cell lineage or as fusion tags to monitor
XX XX protein localisation within living cells. The engineered fluorescent
XX XX proteins are particularly useful for coupling engineered fluorescent
XX XX proteins to antibodies, nucleic acids or other receptors for use in
XX XX detection assays, e.g. immunossays or hybridisation assays. They are
XX XX useful for tracking the movement of proteins in cells or in systems
XX XX for detecting induction of transfection and for the simultaneous
XX XX measurement of two or more processes within cells. Proteins of the
XX XX invention are also useful as fluorescent energy donors or acceptors
XX XX as well as biosensors for detecting antigens. They are also useful in
XX XX fluorescence resonance energy transfer (FRET). The crystal structure
XX XX of the GFP is useful for designing mutants having altered fluorescence
XX XX characteristics which are particularly useful to identify amino acids
XX XX whose substitution alters fluorescent properties of the protein. The
XX XX crystal structure of the GFP is also useful for designing mutants having
XX XX altered antigen binding characteristics which are particularly useful for
XX XX identifying amino acids whose substitution alters the specificity and
XX XX affinity of the binding site to various antigens, and for monitoring antigen
XX XX binding and therefore the concentration of the antigen. The present
XX XX sequence is Aequeorea victoria GFP mutant protein (S65Y).
XX XX Note: This sequence is not shown in the specification but is derived
XX XX from Aequeorea victoria wild type green fluorescent protein shown in

change in fluorescent characteristics. They are useful as markers for gene expression, tracers of cell lineage or as fusion tags to monitor protein localisation within living cells. The engineered fluorescent proteins are particularly useful for coupling engineered fluorescent proteins to antibodies, nucleic acids or other receptors for use in detection assays, e.g. immunoassays or hybridisation assays. They are useful for tracking the movement of proteins in cells or in systems for detecting induction of transcription and for the simultaneous measurement of two or more processes within cells. Proteins of the invention are also useful as fluorescent energy donors or acceptors as well as biosensors for detecting anions. They are also useful in fluorescence resonance energy transfer (FRET). The crystal structure of the GFP is useful for designing mutants having altered fluorescent characteristics which are particularly useful to identify amino acids whose substitution alters fluorescent properties of the protein. The crystal structure of the GFP is also useful for designing mutants having altered union binding characteristics which are particularly useful for identifying amino acids whose substitution alters the specificity and affinity of the binding site to various anions, e.g. for monitoring anion binding and therefore the concentration of the anion. The present sequence is Aquorea victoria GFP mutant protein (A6A1).

Note: This sequence is not shown in the specification but is derived from Aquorea victoria wild type green fluorescent protein shown in figure 5 of the specification (A6A1638).

XX Sequence 238 AA:

Query Match 95.2% Score 12721 DP 211 Length 238
 Best Local Similarity 91.6% Prod No. 9.3e+124
 Matches 237: Conservative 17 Mismatches 17 Gaps 0

OY	1	MSKGEELFGVWFLVVELDGVNGHKEVSUSGSDERDAVYQKILKEICTCKIKVWNTFL	60
DB	1	MSKGEELFGVWFLVVELDGVNGHKEVSUSGSDERDAVYQKILKEICTCKIKVWNTFL	60
OY	61	VTFPSYGVQCFKRYEDHMKKEHFFKSAHPYVQCFTHPEKQKQNYKTAAPVPEGCHLY	120
DB	61	VTFPSYGVQCFKRYEDHMKKEHFFKSAHPYVQCFTHPEKQKQNYKTAAPVPEGCHLY	120
OY	121	NRIELKSLDPRKNNYLDHKLNNKNNNNYDAAKQKNTAUNTSLDRIKFGSSVLA	180
DB	121	NRIELKSLDPRKNNYLDHKLNNKNNNNYDAAKQKNTAUNTSLDRIKFGSSVLA	180
OY	181	HYCQNTPIGDPVLLDNIHVLSTQSALSKDNEKSDHIVVLEFFVTAAGTQMTFLYK	240
DB	181	HYCQNTPIGDPVLLDNIHVLSTQSALSKDNEKSDHIVVLEFFVTAAGTQMTFLYK	240

Search completed: July 16, 2003, 16:10:36
 Job time : 47 secs

Sendcore version 5.1.1
Copyright (c) 1993 - 2003 Computer Corp.

on protein - protein search, using sw model

Run on: July 16, 2003, 16:00:47 Search time 45 seconds

(without alignment)
204,736 Million bits of data/300

Title: US-09-967-301-3

Perfect score: 1256
Sequence: 1 MSKGFRTGVVPIVVELDQ.....VLDVFVTAATVETRMFLYK 214

Scoring table: BLOSUM62
Gapop 10.0, Gapext 1.5

Seatched: 924470 seqs, 135250420 residues

Total number of hits satisfying chosen parameters: 924470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcdqata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcdqata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcdqata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcdqata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcdqata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcdqata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcdqata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcdqata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcdqata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcdqata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcdqata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcdqata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcdqata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcdqata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcdqata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcdqata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcdqata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcdqata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcdqata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcdqata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcdqata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcdqata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcdqata/geneseq/geneseq-emb1/AA2002.DAT.*

Fold. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	db	20	Description
1	1266	99.2	238	23	AAE15087	Aequorea victoria
2	1255	98.7	238	19	AAW76106	A. victoria green
3	1259	98.7	238	23	AAE16073	Aequorea victoria
4	1258	98.6	238	17	AAW5304	Green fluorescent
5	1258	98.6	238	18	AAW24242	Aequorea victoria
6	1258	98.6	238	19	AAW76371	A. victoria green
7	1258	98.6	238	19	AAW76105	A. victoria green
8	1258	98.6	238	19	AAW40479	A. victoria green
9	1258	98.6	238	22	AAW12552	Wild-type green fl
10	1259	98.6	238	23	AAE16036	Aequorea victoria

1	1257	96.5	232	19	AAW65084	A. victoria green
2	1255	96.4	238	19	AAW22101	Aequorea victoria
3	1255	96.4	238	19	AAW76109	A. victoria green
4	1255	96.4	238	19	AAW76110	A. victoria green
5	1255	96.4	238	19	AAW76111	A. victoria green
6	1255	96.4	238	19	AAW65078	A. victoria green
7	1255	96.4	238	19	AAW65079	A. victoria green
8	1255	96.4	238	20	AAW66228	Humanised green fl
9	1255	96.4	238	21	AAW93976	Synthetic green fl
10	1255	96.4	238	23	AAW6042	Aequorea victoria
11	1255	96.4	238	23	AAW6043	Aequorea victoria
12	1255	96.4	238	23	AAW6044	Humanised green fl
13	1255	96.4	238	20	AAW6045	Humanised green fl
14	1255	96.4	238	19	AAW6046	Humanised green fl
15	1255	96.4	238	19	AAW6047	Humanised green fl
16	1255	96.4	238	19	AAW6048	Humanised green fl
17	1255	96.4	238	19	AAW6049	Humanised green fl
18	1255	96.4	238	19	AAW6050	Humanised green fl
19	1255	96.4	238	19	AAW6051	Humanised green fl
20	1255	96.4	238	19	AAW6052	Humanised green fl
21	1255	96.4	238	19	AAW6053	Humanised green fl
22	1255	96.4	238	19	AAW6054	Humanised green fl
23	1255	96.4	238	19	AAW6055	Humanised green fl
24	1255	96.4	238	19	AAW6056	Humanised green fl
25	1255	96.4	238	19	AAW6057	Humanised green fl
26	1255	96.4	238	19	AAW6058	Humanised green fl
27	1255	96.4	238	19	AAW6059	Humanised green fl
28	1255	96.4	238	19	AAW6060	Humanised green fl
29	1255	96.4	238	19	AAW6061	Humanised green fl
30	1255	96.4	238	19	AAW6062	Humanised green fl
31	1255	96.4	238	19	AAW6063	Humanised green fl
32	1255	96.4	238	19	AAW6064	Humanised green fl
33	1255	96.4	238	19	AAW6065	Humanised green fl
34	1255	96.4	238	19	AAW6066	Humanised green fl
35	1255	96.4	238	19	AAW6067	Humanised green fl
36	1255	96.4	238	19	AAW6068	Humanised green fl
37	1255	96.4	238	19	AAW6069	Humanised green fl
38	1255	96.4	238	19	AAW6070	Humanised green fl
39	1255	96.4	238	19	AAW6071	Humanised green fl
40	1255	96.4	238	19	AAW6072	Humanised green fl
41	1255	96.4	238	19	AAW6073	Humanised green fl
42	1255	96.4	238	19	AAW6074	Humanised green fl
43	1255	96.4	238	19	AAW6075	Humanised green fl
44	1255	96.4	238	19	AAW6076	Humanised green fl
45	1255	96.4	238	19	AAW6077	Humanised green fl

ALIGNMENTS

RESULT 1	
AAE16087	
ID	AAE16087 standard: Protein: 238 AA.
XX	
AC	AAE16087
XX	
XX	26-MAY-2002 (first entry)
XX	
DE	Aequorea victoria GFP mutant protein (B2220).
XX	
KW	Green fluorescent protein:: cell linkage tracer:: protein localisation::
KW	GFP:: fusion tag:: gene expression marker:: fluorescent energy acceptor::
KW	immunoassay:: hybridisation assay:: fluorescent energy donor:: biosensor::
KW	FRET:: fluorescence resonance energy transfer:: mutant:: mutagen::
XX	
OS	Aequorea victoria.
OS	Synthetic.
XX	
XX	Key
XX	Location/Qualifiers
XX	
XX	Misc-difference 222
XX	/note- "Wild type Glu substituted with Gly"
XX	
XX	WC200109147-A2.
XX	
XX	29-NOV-2001.
XX	
XX	17-MAY-2001; 2001W0-US16149.
XX	
XX	19-MAY-2000; 2000US-0575847.
XX	
XX	(UYOR-) UNIV OREGON STATE.
XX	
XX	Wachter R, Remington SJ;

CC intensities of emission make these products useful in a wide variety of
 CC contexts, such as tracking of differential gene expression and protein
 CC localisation. The mutations can also be created to modify the encoded
 CC GFP so that it only possesses one absorbance peak.

XX
 SQ Sequence 236 AA:

Query Match 98.6%; Score 1254; DB 17; Length 236;
 Best Local Similarity 98.7%; Pred. No. 20-123;
 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLIVELDGVNGHKFSVSGEGSDATYGGKTLKFKCTTGKLPVWPPTL 60
 DB 1 MSKGEELFTGVVPLIVELDGVNGHKFSVSGEGSDATYGGKTLKFKCTTGKLPVWPPTL 60
 QY 61 VTISYGVQCFSRYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEGGTLY 120
 DB 61 VTISYGVQCFSRYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEGGTLY 120
 QY 62 VTISYGVQCFSRYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEGGTLY 120
 DB 62 VTISYGVQCFSRYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEGGTLY 120
 QY 121 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DB 121 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 QY 122 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DB 122 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 QY 123 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DB 123 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 QY 124 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DB 124 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 QY 125 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DB 125 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180

RESULT 5
 AAW24232
 ID AAW24232 standard: protein; 236 AA.

XX AAW24232;
 AC AAW24232;
 DI 02-FEB-1996 (first entry)
 XX Aequorea victoria green fluorescent protein (GFP);
 DE North West Pacific jellyfish: green fluorescent protein (GFP);
 KW fluorescent, resonance energy transfer; FRET; enzymatic assays;
 KW enzyme inhibitor; enzyme promoter; tandem fluorescent protein constructs;
 KW blue fluorescent protein; protease cleavage site.

XX Aequorea victoria.
 CS W09724261-A1.
 XX 07-AUG-1997.
 XX 31-JAN-1997; 97WG-US01457.
 XX 31-JAN-1996; 96US-0594575.
 PR (AURO-) Aequorea BIOSCIENCES CORP.
 PA (REGC) UNIV CALIFORNIA.
 XX
 XX Cubitt A., Heim R., Tsien RY.
 PI WPI: 1997-402615/37.
 DR N-PSDB; AA:77843.
 DR
 DR
 PT Tandem fluorescent protein constructs have donor and acceptor
 PT moieties exhibiting fluorescent linked via cleavable peptide linker.
 PI useful in enzymatic assays

XX
 XX
 PS Disclosure: Fig 1; 80pp; English.

CC This sequence represents Aequorea victoria (North West Pacific jellyfish);
 CC green fluorescent protein (GFP). The GFP protein is used in a novel
 CC tandem fluorescent protein construct. The construct comprises a donor
 CC (e.g. GFP protein) and an acceptor (e.g. S65C, a variant of GFP or
 CC blue fluorescent protein) fluorescent protein moiety, and a linker
 CC coupling them. Preferably, the donor is positioned at the N-terminus of

CC the polypeptide relative to the acceptor. The linker moiety is a peptide
 CC 5-50 amino acids in length containing a protease cleavage site. The donor
 CC and acceptor moieties exhibit fluorescent resonance energy transfer
 CC (FRET) when the donor moiety is excited. The constructs are used in
 CC enzymatic assays and can be used to isolate new enzymes or enzyme
 CC inhibitors or promoters. The specific activity of enzyme (in vivo and in
 CC vitro) and compounds altering enzyme activity can be obtained. FRET and
 CC hence activity of specific compounds is measured from the acceptor or
 CC donor moiety or maybe obtained using a ratio between the two.

XX
 SQ Sequence 238 AA:

Query Match 98.6%; Score 1258; DB 18; Length 238;
 Best Local Similarity 98.7%; Pred. No. 20-123;
 Matches 235; Conservative 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLIVELDGVNGHKFSVSGEGSDATYGGKTLKFKCTTGKLPVWPPTL 60
 DE 1 MSKGEELFTGVVPLIVELDGVNGHKFSVSGEGSDATYGGKTLKFKCTTGKLPVWPPTL 60
 QY 61 VTISYGVQCFSRYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEGGTLY 120
 DE 61 VTISYGVQCFSRYPDHMKRHDFFKSAMPEGVQVQERTIFFKDDGNYKTRAEVKEGGTLY 120
 QY 121 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DE 121 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 QY 122 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DE 122 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 QY 123 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DE 123 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 QY 124 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DE 124 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 QY 125 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180
 DE 125 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVFNKTRAEVKEGGTLY 180

RESULT 6
 AAW76371
 ID AAW76371 standard: protein; 238 AA.

XX AAW76371;
 AC AAW76371;
 DI 03-DEC-1998 (first entry)
 XX A. victoria green fluorescent protein.
 DE Green fluorescent protein: GFP; jellyfish: detection; mutational activity;
 KW distribution.

XX Aequorea victoria.
 CS J010234382-A.
 XX 08-SEP-1998.
 PD 27-FEB-1997; 97JP-0062370.
 XX 27-FEB-1997; 97JP-0062370.
 XX 27-FEB-1997; 97JP-0062370.
 XX (CINA) DINABEKKE KENYUSHO KK.
 PA WPI: 1998-515036/46.
 DR N-PSDB; AAW61751.
 DR
 DR
 PT Green fluorescent protein: useful for determining promoter and
 PT targeting activity in a cell

XX
 XX
 PS Disclosure: Page 6-7; 11pp; Japanese.

CC This sequence represents a wild-type green fluorescent protein, GFP,
 CC isolated from Aequorea victoria. This protein is used in the
 CC construction of a mutant protein which emits fluorescence 6 times
 CC higher than the conventional fluorescent protein. The nuclear acid can
 CC be used for the recombinant production of the fluorescent protein and
 CC also for the determination of activity of a promoter to be tested,
 CC comprising steps of introducing the above vector to a host cell and

CC detecting the fluorescence emitted from the cell, they can additionally
CC be used for detecting the targeting activity in the cell of an amino
CC acid sequence to be tested in which the above fluorescent protein is
CC introduced to a cell and the distribution of the fluorescent protein in
CC the cell is observed.

XX Sequence 238 AA:
Query Match: 98.6%, Score: 1254, DB: 19, Length: 238;
Best Local Similarity: 96.7%, Pred. No. 2e-123;
Matches: 235; Conservative: 0; Mismatches: 3; Indels: 0; Gaps: 0;
QY 1 MSKGEELFTGVPLVILVGLDGVNKHKFSVSGEGGATATKTLKFTCTTKSLFVPEWTL 60
DB 1 MSKGEELFTGVPLVILVGLDGVNKHKFSVSGEGGATATKTLKFTCTTKSLFVPEWTL 60
QY 61 VTLISYGVGCFSRYPDHNKREDFKSAKPGYVCEITFFKUCGNGKTRAEVKEGUILY 120
DB 61 VTLISYGVGCFSRYPDHNKREDFKSAKPGYVCEITFFKUCGNGKTRAEVKEGUILY 120
QY 121 NRLEKGIIDKEDNLTGKLEYNNSHNHYIMADKGNKATKVNFKIRHKEATVGLAD 180
DB 121 NRLEKGIIDKEDNLTGKLEYNNSHNHYIMADKGNKATKVNFKIRHKEATVGLAD 180
QY 181 HYQONTPTGSGVLPNNHYSTOSALSKEPNEDHNVILGAVIAAGITGGKHELYK 238
DB 181 HYQONTPTGSGVLPNNHYSTOSALSKEPNEDHNVILGAVIAAGITGGKHELYK 238

RESULT 7
AAW76105
ID AAW76105 standard; Protein: 238 AA.
AC AAW76105;
XX 18-NOV-1998 (first entry)
DE A. victoria green fluorescent protein.
KW Green fluorescent protein; GFP; mutant; yellowish excitation;
KW chromophore.
XX Aequorea victoria.
XX US8804357-A. 97US-0791342.
XX 31-JAN-1997;
XX 01-FEB-1998; 96US-0019560.
XX 31-JAN-1997; 97US-0791342.
XX (STR) UNIV LELAND STANFORD JUNIOR.
XX Cornack BP, Falkow S, Vardivola RH;
XX WPI: 1998-505643/42.
XX DNA encoding mutant green fluorescent protein. With
XX greater fluorescence intensity than wild type proteins, useful for
XX studying gene expression and protein localization.
XX Disclosure: Page 1; 15pp; English.

XX This sequence represents the green fluorescent protein (GFP) from
XX Aequorea victoria. This sequence is used in a method to discover GFPs
XX mutated in the chromophore region that fluoresce more brightly than
XX wild-type GFP upon excitation at 488 nm. These mutants can be used in a
XX method for analyzing a cell containing the GFP. The GFP can also be fused
XX to a protein of interest and used to identify the intracellular localization of a
XX protein of interest. A regulatory element could be operatively connected
XX to a coding portion encoding a mutant GFP and exposed to an environmental

CC stimulus. The fluorescence signal from the from the cell then measures
CC the effect of the stimulus on the regulatory element.
CC Note: This sequence does not appear in the specification but has
CC been retrieved from the EMBL record contained in W05134.

XX Sequence 238 AA:
Query Match: 98.6%, Score: 1254, DB: 19, Length: 238;
Best Local Similarity: 96.7%, Pred. No. 2e-123;
Matches: 235; Conservative: 0; Mismatches: 3; Indels: 0; Gaps: 0;
QY 1 MSKGEELFTGVPLVILVGLDGVNKHKFSVSGEGGATATKTLKFTCTTKSLFVPEWTL 60
DB 1 MSKGEELFTGVPLVILVGLDGVNKHKFSVSGEGGATATKTLKFTCTTKSLFVPEWTL 60
QY 61 VTLISYGVGCFSRYPDHNKREDFKSAKPGYVCEITFFKUCGNGKTRAEVKEGUILY 120
DB 61 VTLISYGVGCFSRYPDHNKREDFKSAKPGYVCEITFFKUCGNGKTRAEVKEGUILY 120
QY 121 NRLEKGIIDKEDNLTGKLEYNNSHNHYIMADKGNKATKVNFKIRHKEATVGLAD 180
DB 121 NRLEKGIIDKEDNLTGKLEYNNSHNHYIMADKGNKATKVNFKIRHKEATVGLAD 180
QY 181 HYQONTPTGSGVLPNNHYSTOSALSKEPNEDHNVILGAVIAAGITGGKHELYK 238
DB 181 HYQONTPTGSGVLPNNHYSTOSALSKEPNEDHNVILGAVIAAGITGGKHELYK 238

RESULT 8
AAW40475
ID AAW40475 standard; Protein: 238 AA.
AC AAW40475;
XX 15-SEP-1998 (first entry)
DE A. victoria green fluorescent protein.
KW Green fluorescent protein; protein kinase; phosphate donor;
KW phosphorylation; drug screening; receptor-ligand binding; signaling;
KW protein-protein interaction; kinase activation.
XX Aequorea victoria.
XX Key: Location/Qualifiers
XX Protein: 1..238
XX /cont= "partial sequence"
XX W09632571-A1.
XX 22-JAN-1998;
XX 16-JUL-1997; 97W-0512410.
XX 16-JUL-1998; 98US-0640873.
XX 16-JUL-1998; 98US-0679865.
XX 16-JUL-1998; 98US-0630876.
XX (PDB) UNIV CALIFORNIA.
XX CDBIT: A5, Tsien RY;
XX WPI: 1988-110416/17;
XX N-PSDB: AAW11125.
XX Assays for protein kinase and modulators - using a fluorescent
XX protein substrate which exhibits a different fluorescent property in
XX the phosphorylated and in-phosphorylated state
XX Disclosure: Fig 1; 65pp; English.
XX This sequence represents a green fluorescent protein from Aequorea
XX victoria which is used in a method to determine if a sample contains

XX (JYOR-) UNIV GREGG STATE
XX
XX
XX Wachter R. Remington SJ
XX
XX WPI: 2002-093064/11.
XX N-PSDB: AAG26107.
XX
XX New long wavelength engineered fluorescent proteins, useful as markers
XX for gene expression, tracers of cell lineage or as fusion tags to
XX monitor protein localization, or in detection assays, cell immunoassays
XX or hybridization assays
XX
XX Claim 27: Fig 3: 1a1pp: English.
XX
XX The patent discloses functional engineered fluorescent protein and the
XX corresponding polynucleotide. The amino acid sequence of the engineered
XX protein is identical to Acetone green fluorescent protein (GFP). The
XX engineered fluorescent proteins of the invention have varied fluorescent
XX properties and have the ability to respond to ion concentrations via a
XX change in fluorescent characteristics. They are useful as markers for
XX gene expression, tracers of cell lineage or as fusion tags to monitor
XX protein localization within living cells. The engineered fluorescent
XX proteins are particularly useful for coupling engineered fluorescent
XX proteins to antibodies, nucleic acids or other receptors for use in
XX detection assays, e.g. immunoassays or hybridization assays. They are
XX useful for tracking the movement of proteins in cells or in systems
XX for detecting induction of transcription and for the simultaneous
XX measurement of two or more processes within cells. Proteins of the
XX invention are also useful as fluorescent energy donors or acceptors
XX as well as biosensors for detecting anions. They are also useful in
XX fluorescence resonance energy transfer (FRET). The crystal structure
XX of the GFP is useful for designing mutants having altered fluorescent
XX characteristics which are particularly useful to identify amino acids
XX whose substitution alters fluorescent properties of the protein. The
XX crystal structure of the GFP is also useful for designing mutants having
XX altered anion binding characteristics which are particularly useful for
XX identifying amino acids whose substitution alters the specificity and
XX affinity of the binding site to various anions, and for monitoring anion
XX binding and therefore the concentration of the anion. The present
XX sequence is Acetone victoria green fluorescent protein.
SQ Sequence 238 AA:
Query Match 98.6% Score 1259; ID 23; Length 238;
Best Local Similarity 98.7% Pred No 20-123;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPLVLELSDYNGHKGKFSVSGEGDGYGKLTKEFICTGKLPVWPTL 60
DE 1 MSKGEELFTGVVPLVLELSDYNGHKGKFSVSGEGDGYGKLTKEFICTGKLPVWPTL 60
QY 61 VTTLGVGOCFSRYPUHMKRHDFEFKSMPEGYVQERTIFFKDGMYKTRAEVKGFTGLV 120
DE 61 VTTLGVGOCFSRYPUHMKRHDFEFKSMPEGYVQERTIFFKDGMYKTRAEVKGFTGLV 120
QY 121 NRLEKGTDEKESNLTHRIENYNNSHWYVWACKGKHKYNKRIKHNIEGKQGLAD 180
DE 121 NRLEKGTDEKESNLTHRIENYNNSHWYVWACKGKHKYNKRIKHNIEGKQGLAD 180
QY 181 HYQNTDPIGPGVLLPNHHVLTQSALSKDNKRNKRNHVLQFTVAGLTHGHMFLYK 240
DE 181 HYQNTDPIGPGVLLPNHHVLTQSALSKDNKRNKRNHVLQFTVAGLTHGHMFLYK 240
PB 161 HYQNTDPIGPGVLLPNHHVLTQSALSKDNKRNKRNHVLQFTVAGLTHGHMFLYK 240

RESULT 11

AAW65064 standard; Protein; 238 AA.

XX

XX AAW65064;

XX 15-SEP-1996 (first entry)

XX

DE A. victoria green fluorescent protein variant 517K.
XX Green fluorescent protein; protein kinase; phosphate donor;
XX phosphorylation; drug screening; receptor-liquid binding; signalling;
XX protein-protein interaction; kinase activation.
XX Acquired victoria.
XX Synthetic.
XX Key Location/Qualifiers
XX Misc-difference 572 /Label: 517K
XX 909802871-AL. /note= "wild type 618 is replaced with lys. The
XX 22-JAN-1998. specification states the mutation occurs at
XX 16-JUL-1996; 96JS-0680877. position 171 however it is more likely to be at
XX 16-JUL-1996; 96JS-0680876. position 172."
XX (580C) UNIV CALIFORNIA.
XX CUBANA AB. Lisbon KY;
XX WPI: 1998-110615/10.
XX Assays for protein kinase and modulators - using a fluorescent
XX protein substrate which exhibits a different fluorescent property in
XX the phosphorylated and un-phosphorylated state
XX Claim 35; Page 1: 65pp: English.
XX This sequence represents a green fluorescent protein variant from
XX Acetone victoria in which a wild-type Gln residue at position 172 is
XX replaced with a lys. The specification states that this mutation occurs
XX at position 171 however the wild type Gln residue is found at position
XX 172. This variant is used in a method to determine if a sample contains
XX protein kinase (PK) activity. The method involves contacting the sample
XX with a phosphate donor (GDP) and a fluorescent protein (FP) substrate for
XX a PK. the protein substrate comprising a FP moiety and a phosphorylation
XX site for a PK, where the protein substrate exhibits a different
XX fluorescent property in the phosphorylated state than in the
XX un-phosphorylated state. The protein substrate is then excited and the
XX amount of a fluorescent property that differs in the un-phosphorylated
XX state and phosphorylated state is measured, whereby an amount that is
XX consistent with the presence of the protein substrate in its
XX phosphorylated state indicates the presence of PK activity. The method
XX and products can be used in drug screening. They can be used for
XX screening for compounds which affect cellular events, including
XX receptor-liquid binding, protein-protein interactions or kinase
XX activation, which signal to the target kinase.
XX NOTE: this sequence does not appear in the specification and has
XX been constructed from the wild-type sequence represented in AAW6479.
SQ Sequence 238 AA:
Query Match 98.6% Score 1257; ID 19; Length 238;
Best Local Similarity 98.6% Pred No 2-5e-123;
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPLVLELSDYNGHKGKFSVSGEGDGYGKLTKEFICTGKLPVWPTL 60
DE 1 MSKGEELFTGVVPLVLELSDYNGHKGKFSVSGEGDGYGKLTKEFICTGKLPVWPTL 60
QY 61 VTTLGVGOCFSRYPUHMKRHDFEFKSMPEGYVQERTIFFKDGMYKTRAEVKGFTGLV 120
DE 61 VTTLGVGOCFSRYPUHMKRHDFEFKSMPEGYVQERTIFFKDGMYKTRAEVKGFTGLV 120

DB 61 VTTISYGVQCFSRYPDHRKEDFKSAMPEGVYQERTIFPKDGNKYKTRAEVKEFEDTIV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKVNFKRHNIEGSGVQLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKVNFKRHNIEGSGVQLAD 180
 QY 181 HYQONTPIGSGPVLIPDNHYLSTQSALSKDPNKRKHVLLSEFVTAAGITHEGMDELYK 238
 DB 181 HYQONTPIGSGPVLIPDNHYLSTQSALSKDPNKRKHVLLSEFVTAAGITHEGMDELYK 238

RESULT 12
 ID AAW22101 standard; protein: 238 AA.
 AC AAW22101:
 DT 10-MAR-1998 (first entry)
 DE Aequorea victoria protein GFP10 mutant F64L-S65L.
 XX Green fluorescent protein; GFP; jellyfish; GFP10 gene; reporter;
 KW humaniser; expression; mutant.
 XX Aequorea victoria.
 OS
 XX KEY Location/Qualifiers
 FE Misc-difference 64 /Label= F64L
 FI /note= "Wild-type Phe is replaced by Ier"
 FT Misc-difference 65 /Label= S65L
 FE /note= "Wild-type Ser is replaced by Thr"
 FT
 XX W09726333 Ala.
 XX 24-JUL-1997.
 XX 17-JAN-1997: 97WC-000755.
 XX 18-JAN-1996: 96JS-0586201.
 XX (GFFL) UNIV FLORIDA RES FOUND INC.
 XX Hauswirth WM, Rzyzcka N, Zolotukhin S;
 XX W21: 1997-383337/45.
 XX Humanised green fluorescent protein gene - optimised to provide high
 PT level expression in mammalian cells, used e.g. to label or identify
 FT cells, to locate proteins, etc.
 XX
 XX Claim 115; Page 7; 15pp; English.
 XX This sequence represents a mutant form of the green fluorescent protein
 CC GFP10 i.e. F64L-S65L. Green fluorescent proteins (GFP's) can be used
 CC as reporter molecules as they absorb blue light and emit green light
 CC without requiring any cofactors, substrates, or additional gene products
 CC allowing GFP detection in living cells provided; meaningful gene
 CC expression is achieved. By providing humanised GFP, adapted for
 CC expression in mammalian and human cells, problems associated with
 CC wild-type jellyfish GFP e.g. variable and low expression levels should be
 CC overcome.
 CC Note: The present sequence does not appear in the specification. It
 CC has been made by modifying the wild-type GFP10 amino acid sequence found
 CC in AAW31295.
 XX
 XX Sequence 238 AA:
 SQ

Query Match 98.4%; Score 1355; Pos 12; Length 238;
 Best Local Similarity 98.3%; Pred. No. 4, 1e-124;
 Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGWPIILVELDGVNKHKPSVSGEGEDATYCKLTLKELCTGKLPVWITL 60
 DB 1 MSKGEELFTGWPIILVELDGVNKHKPSVSGEGEDATYCKLTLKELCTGKLPVWITL 60
 QY 61 VTTISYGVQCFSRYPDHRKEDFKSAMPEGVYQERTIFPKDGNKYKTRAEVKEFEDTIV 120
 DB 61 VTTISYGVQCFSRYPDHRKEDFKSAMPEGVYQERTIFPKDGNKYKTRAEVKEFEDTIV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKVNFKRHNIEGSGVQLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKVNFKRHNIEGSGVQLAD 180
 QY 181 HYQONTPIGSGPVLIPDNHYLSTQSALSKDPNKRKHVLLSEFVTAAGITHEGMDELYK 238
 DB 181 HYQONTPIGSGPVLIPDNHYLSTQSALSKDPNKRKHVLLSEFVTAAGITHEGMDELYK 238

RESULT 13
 ID AAW76109 standard; protein: 238 AA.
 AC AAW76109:
 DT 18-NOV-1998 (first entry)
 DE A. victoria green fluorescent protein mutant S72A.
 XX Green fluorescent protein; GFP; mutant; jellyfish; excitation;
 KW chromophore.
 XX Acquired victoria.
 OS Synthetic.
 XX KEY Location/Qualifiers
 FE Misc-difference 72 /Label= S72A
 FT /note= "Wild-type Ser is replaced by Ala"
 XX US5504367-A.
 XX 05-SEP-1998.
 XX 31-JAN-1997: 97JS-0791342.
 XX 31-SEP-1996: 96JS-0070960.
 XX 31-JAN-1997: 97JS-0791342.
 XX (SIR) UNIV LELAND STANFORD JUNIOR.
 XX Cornack BP, Falkow S, Valdivia PH;
 XX W21: 1998-505643/43.
 XX DNA encoding mutant green fluorescent pigment proteins - with
 FT greater fluorescence intensity than wild-type proteins, useful for
 PT studying gene expression and protein localisation
 XX
 XX Claim 5; Page 7; 15pp; English.
 XX This sequence represents a mutant green fluorescent protein (GFP) from
 CC Aequorea victoria in which a Ser residue at position 72 in the wild
 CC type protein is replaced by an Ala. These mutations occur in the
 CC chromophore region. GFP mutants are used in a method to discover GFP's
 CC mutated in the chromophore region that fluoresce more brightly than
 CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a
 CC method for analysing a cell containing the GFP. The GFP can also be
 CC fused to a protein and used to identify the intracellular localisation of
 CC a protein of interest. A regulatory element could be operative, if
 CC connected to a coding portion encoding a mutant GFP and exposed to an
 CC environmental stimulus. The fluorescence signal from the cell
 CC then measures the effect of the stimulus on the regulatory element.
 CC NOTE: This sequence does not appear in the specification but has been
 CC constructed from the wild-type GFP protein represented in AAW76106

```

XX
SQ Sequence 238 AA:
Query Match 98.4%, Score 1255, Pos 19, Length 238:
Best Local Similarity 98.38, Pred. No. 4, 10-123:
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0
QY 1 MSKGEELPTGVVFLVILVGGVNGHKFSVSQSGRQDAVYGLKLEKFIETGKLVVWPTL 60
Dc 1 MSKGEELPTGVVFLVILVGGVNGHKFSVSQSGRQDAVYGLKLEKFIETGKLVVWPTL 60
QY 61 VTTLSVYVQCFSPRYPOBKHKHDFKFSAMFESYVOERTIFFKDDGNYKTRAFKPESTLV 120
Dc 61 VTTLSVYVQCFSPRYPOBKHKHDFKFSAMFESYVOERTIFFKDDGNYKTRAFKPESTLV 120
QY 121 NRLEKGGTDEKEDGNLGHKLEYNNSNNVIMADKKNKGIKVNFKIRHNLEAGVOLAD 180
Dc 121 NRLEKGGTDEKEDGNLGHKLEYNNSNNVIMADKKNKGIKVNFKIRHNLEAGVOLAD 180
QY 181 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHNVLLGCVTAARITHQREDEYK 240
Dc 181 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHNVLLGCVTAARITHQREDEYK 240
Dc 191 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHNVLLGCVTAARITHQREDEYK 240

RESULT 14
AAW76110
ID AAW76110 standard: protein: 238 AA.
AC AAW76110:
DT 18-NOV-1998 (first entry)
XX A. victoria green fluorescent protein mutant: S65A.
XX Green fluorescent protein: GFP; mutant: jellyfish; excitation:
XX chromophore.
OS Aequorea victoria.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 65 /label- S65A
FT /note- "Wild-type Ser is replaced by Ala"
PN US5864387-A.
PC 08-SEP-1994.
XX 31-JAN-1997: 97US-0791332.
XX 01-FEB-1996: 96US-0610467.
XX 31-JAN-1997: 97US-0791332.
XX (STRD ) UNIV LELAND STANFORD JUNIOR.
XX Cormack RP, Falkow S, Valdivia RH.
XX WPI: 1998-505643/43.
XX DNA encoding mutant green fluorescent protein - with
XX greater fluorescence intensity than wild-type protein. useful for
XX studying gene expression and protein localisation
XX Claim 6; Page 1; 15pp; English.
XX This sequence represents a mutant green fluorescent protein (GFP) from
XX Aequorea victoria in which a Ser residue at position 65 in the wild
XX type protein is replaced by an Ala. These mutations occur in the
XX chromophore region. GFP mutants are used in a method to discover GFP's
XX mutated in the chromophore region that fluoresce more brightly than
XX wild-type GFP upon excitation at 488 nm. These mutants can be used in a
XX method for analyzing a cell containing the GFP. The GFP can also be

```

```

CC fused to a protein and used to identify the intracellular localisation of
CC a protein of interest. A regulatory element could be operatively
CC connected to a coding portion encoding a mutant GFP and exposed to an
CC environmental stimulus. The fluorescence signal from the cell
CC then measures the effect of the stimulus on the regulatory element.
CC NOTE: This sequence does not appear in the specification but has been
CC constructed from the wild-type GFP protein represented in AAW71106
SQ Sequence 238 AA:
Query Match 98.4%, Score 1255, Pos 19, Length 238:
Best Local Similarity 98.38, Pred. No. 4, 10-123:
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0
QY 1 MSKGEELPTGVVFLVILVGGVNGHKFSVSQSGRQDAVYGLKLEKFIETGKLVVWPTL 60
Dc 1 MSKGEELPTGVVFLVILVGGVNGHKFSVSQSGRQDAVYGLKLEKFIETGKLVVWPTL 60
QY 61 VTTLSVYVQCFSPRYPOBKHKHDFKFSAMFESYVOERTIFFKDDGNYKTRAFKPESTLV 120
Dc 61 VTTLSVYVQCFSPRYPOBKHKHDFKFSAMFESYVOERTIFFKDDGNYKTRAFKPESTLV 120
QY 121 NRLEKGGTDEKEDGNLGHKLEYNNSNNVIMADKKNKGIKVNFKIRHNLEAGVOLAD 180
Dc 121 NRLEKGGTDEKEDGNLGHKLEYNNSNNVIMADKKNKGIKVNFKIRHNLEAGVOLAD 180
QY 181 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHNVLLGCVTAARITHQREDEYK 240
Dc 181 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHNVLLGCVTAARITHQREDEYK 240
Dc 191 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHNVLLGCVTAARITHQREDEYK 240

RESULT 15
AAW76111
ID AAW76111 standard: protein: 238 AA.
AC AAW76111:
DT 18-NOV-1998 (first entry)
XX A. victoria green fluorescent protein mutant V68L.
XX Green fluorescent protein: GFP; mutant: jellyfish; excitation:
XX chromophore.
OS Aequorea victoria.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 68 /label- V68L
FT /note- "Wild-type Val is replaced by Leu"
PN US5864387-A.
PC 08-SEP-1994.
XX 31-JAN-1997: 97US-0791332.
XX 01-FEB-1996: 96US-0610460.
XX 31-JAN-1997: 97US-0791332.
XX (STRD ) UNIV LELAND STANFORD JUNIOR.
XX Cormack RP, Falkow S, Valdivia RH.
XX WPI: 1998-505643/43.
XX DNA encoding mutant green fluorescent protein - with
XX greater fluorescence intensity than wild-type proteins. useful for
XX studying gene expression and protein localisation
XX Claim 7; Page 1; 15pp; English.
XX

```

